Linguistic Diversity and Out-Group Discrimination in Bilingual Societies

María Paz Espinosa [1]
University of the Basque Country (UPV/EHU)

Enrique Fatás
Loughborough University and University of Pennsylvania

Paloma Ubeda
UNED

June, 2018

Abstract

We study ethno-linguistic diversity in a lab-in-the-field experiment in two bilingual societies, with (Bilbao, in the North of Spain) and without group conflict (Valencia, in the East). Participants from two ethno-linguistic cultures interact with other participants in Homogeneous (no diversity) or Mixed (ethno-linguistic diversity) environments. Participants allocate resources in a nested public goods game with a local and a global public good. By design, by not contributing to the global (and efficient) public good participants exclude the out- group from the benefits of their contribution. Our results show that while diversity reduces contributions to the global public good in Bilbao, it does not have a negative effect in Valencia. We interpret the results with a second experiment eliciting social norms, with different subjects, and find that the positive (negative) effects of diversity in Valencia (Bilbao) are linked to optimistic empirical expectations (normative discrimination). We also show how the prevalence of discriminatory norms is sensitive to additional information about behavior.

Keywords: natural identity, ethno-linguistic groups, discrimination, lab-in-the-field experiment, group conflict.
JEL: C92, D74, H45, J15.

[1] *Corresponding author*. University of the Basque Country. Department of Economics. Avenida Lehendakarri Aguirre 83, 48015, Bilbao, Spain. mariapaz.espinosa@ehu.eus

1. Motivation

The economic implications of diversity are a matter of growing importance. While welfare losses associated to the under provision of public goods in Mixed societies have been empirically documented (Alesina et al., 2003), there is still a lack of understanding regarding the behavioral channels for this negative relationship. This paper explores the connection between ethno-linguistic diversity and economic performance through their effect on out-group discrimination in a experiment.

The detrimental consequences of diversity on economic performance include failure to undertake joint projects, lack of cooperation between groups, or missing the opportunity to exploit complementarities. In the extreme case, armed conflicts may generate huge economic losses due to the destruction of assets or arms investment (Neary, 1997).[2] Ethno-linguistic diversity may harm economic activity by reducing growth rates, investment and the quality of government, and limiting the provision of public goods (Easterly and Levine, 1997; Banerjee et al., 2005; La Porta et al., 1999; Alesina et al., 2003; Alesina and La Ferrara, 2005; Alesina and Zhuravskaya, 2011; Li, 2010, among others). In their study of ethnic diversity and local public goods in rural Western Kenya, Miguel and Gugerty (2005) conclude that ethnic diversity is associated with lower primary school funding and poorer school facilities.

Even after controlling for other socioeconomic and demographic variables, Alesina et al., (1999) document how spending on productive public goods in US cities is inversely related to their ethnic fragmentation. Interestingly, they also find that voters prefer to reduce expenditure on public goods when shared with other ethnic groups. Using cross section data on 225 countries, Desmet, et al. (2009) find that linguistic diversity within a country lowers redistribution,[3] and Khwaja (2009) shows that social fragmentation has similar negative consequences for the upkeep of local public goods in Northern Pakistan. Ethnic diversity is identified in all these studies as a possible source of under provision of public goods, and inefficiency.

---

[2] Armed conflict is an extreme form of inefficiency. Gaibulloev and Sandler (2008, 2009), Abadie (2006) and Abadie and Gardeazabal (2003, 2008) have analyzed the economic consequences of terrorism. We here focus on the cost of non-violent conflict between groups and the mechanisms by which it affects economic relationships, efficiency or social welfare.

[3] Using government transfers and subsidies as a percentage of GDP as a proxy. With an average level of redistribution of 9.5% of GDP in their data set, an increase by one standard deviation in the degree of diversity lowers redistribution by approximately one percentage point.

The strong evidence on the negative consequences of diversity on economic performance typically fails to identify the channels of transmission from ethnic diversity to under provision at the behavioral level. The use of experiments may help to understand some of these channels. As a fast-growing literature in psychology and economics has repeatedly documented, even artificially created groups exhibit different forms of in-group favoritism and out-group discrimination (in the so-called *minimal group paradigm*, starting with Tajfel and Turner, 1979).

A vast behavioral literature has studied how artificially created diversity may have first order effects on economic performance, including within-team diversity in organizational settings (Brandts et al., 2007, 2016), asymmetries in group size (Abbink et al., 2010), endowments (Hargreaves Heap et al., 2015), punishment (Sääksvuori et al.; 2011), punishment and endowment (Weng and Carlsson, 2015), effort costs (Bhattacharya, 2016), and abilities (Chen and Lim, 2017; Brookins et al. 2015). In a seminal paper, Chen and Li (2009) show how group identity shapes social preferences. Interestingly, group membership may also alter the expectations regarding the behavior of in-group and out-group individuals. Therefore, minimal group identity may change economic behavior through its effect on individual social preferences and beliefs.[4]

While artificial diversity may generate lack of cooperation, and the *minimal group paradigm* may capture important underlying behavioral forces, several papers in recent years have used a different approach. Rather than creating simulated identities, these papers exploit the existence of naturally occurring groups in different settings to study the link between diversity and efficiency.[5] Bernhard et al. (2006) use dictator game experiments in two social groups in Papua New Guinea and support the existence of in-group favoritism as exacerbated altruism towards in-group members and more indulgent judgment with in-group norm violators. Tanaka and Camerer (2009) conducted experiments with three different communities in Vietnam (Vietnamese, Chinese and Khmer) finding no in-group favoritism for two of the communities in their interaction with the poor minority and a strong in-group effect in the poor minority.

[4] See also Charness et al. (2007), Abbink et al. (2010), Eckel and Grossman (2005) and Eckel et al. (2010).

[5] Akerlof and Kranton (2000, 2002, 2005) introduced the idea of identity as a variable affecting individuals' economic behavior. See also Aguiar et al. (2010). Social identity is considered multidimensional (see Benabou and Tirole, 2011, for a general model of identity); it may include ethnicity, language, gender, socio-economic level, occupation, etc.

Goette et al. (2006) study subjects randomly assigned to different platoons during a four-week period of officer training in the Swiss Army and find increased willingness to cooperate with fellow platoon members in a sequence of experiments. In a laboratory experiment Keuschnigg and Schikora (2014) find that cultural diversity (Hindu and Muslim subjects in India) does not affect contributions to a static public good game, although it reduces cooperation in the presence of a leader. Our framework is similar in that we also study the effect of cultural diversity in a public good game but we inquire into the behavioral channels that operate; in particular, we analyze the dynamics of conditional cooperation in two different settings and the ethnic groups' first and second order beliefs.

Benjamin et al. (2016) prime the religious identity of participants in a series of one-shot public goods experiments and find that the identity effect strongly varies by religion, with no evidence of religious priming effects on discount rates or generosity. Habyarimana et al. (2007) identify several mechanisms that could explain why ethnic diversity generates low provision of public goods and provide solid experimental evidence about the connection between cooperation in Homogeneous communities in Uganda and in-group norms, supported by expectations that non-contributors will easily face punishment.

Candelo et al. (2017) focus on a fast-growing community in the US, Hispanics, who have a distinct natural identity, to study varied and naturally-occurring individual-level measures of identity and exclusion captured in a post-experimental questionnaire. Participants are recruited in three neighborhoods in Dallas, and may contribute to both a linear public good (in an abstract and artificial environment similar to the one existent in other laboratory studies) and to local organizations providing public goods in their communities. While contributions (in and outside the laboratory) increase with the strength of their self-reported identity, their donations to community charities decrease with their self-reported perception of being socially excluded.

Li et al. (2017) also focus on two low- and middle-income neighborhoods in Dallas to study how naturally occurring community identity has an effect on donations to local organizations. Identity is primed in one treatment and donations compared to a control treatment with no identity priming, controlling for individual perceptions of participants' social image of their communities. While donations (insignificantly)

increase when identity is primed in one of the neighborhoods, they (significantly) decrease in the other. The results are consistent with the existence of very different perceptions of the common identity. While the post-experimental questionnaire reveals that in the former common identity has positive connotations, in the latter has negative ones. The connection is channeled by very different memories of community experiences triggered by identity priming.

There is a vast literature studying intergroup effects using different settings like the dictator game (Ahmed, 2007), the ultimatum game (Kubota et al., 2013, Borstein and Yaniv, 1998), the trust game (Guth et al., 2008, Fershtman and Gneezy, 2001), the prisoners' dilemma (Chuah et al., 2014), and in non-strategic settings (Fatás et al. 2018). While other strategic environments may generate interesting evidence in other domains, our nested public goods game is very well suited to study the natural characteristics of participants in Experiment 1 and 2.

In this paper, we contribute to the literature studying the connection between public good provision and diversity, and document in a laboratory experiment the very different effects of ethnolinguistic diversity in two bilingual societies. We explain how diversity may play a positive or negative effect on performance, studying the different empirical and injunctive norms in both bilingual societies.

In our Experiment 1, we study naturally occurring groups in two bilingual locations in Spain. By keeping the assignment mechanism constant (the self-reported mother tongue language), we manipulate the level of diversity and generate a clear-cut distance between *Homogeneous* settings, when all participants share the same identity and speak the same language outside the laboratory and *Mixed* sessions in which the linguistic identity of the participants is not the same. We run the same experiment in two different locations with very different perceptions of their identities. As the following section documents, and Table A2 supports, in the Basque Country a significant fraction of the local population feel exclusively Basque (21%), or more Basque than Spanish (28%). In Valencia, these two views are hold by a very small minority of individuals (11% and 2%, respectively). We briefly introduce the reader to the cultural and social differences between the two locations in the next section.

Participants in Experiment 1 make an allocation decision between two different public goods and a private good. As described in Section 3, they play a repeated and linear

nested public goods game. In each of the 20 rounds of the game, they distribute their round endowment between a private account, a *local* public good (benefiting only their *local* group) and a *global* public good (benefiting both local groups). As the return of both public goods is the same -they share an identical marginal per capita return-, efficiency is fully driven by group size. As the size of both local groups is the same, fully contributing to the global public good is socially efficient, as it is the way to maximize social welfare (and earnings). By construction, this nested public good game exhibits a useful property: a token contributed to the global public good generates as much return to the other members of the own local group as a token directly contributed to the local public good (because the marginal per capita returns are the same); the only difference being that global contributions benefit the members of the other group as well, while local contributions do not. In other words, when the allocation is *local*, out-group members are automatically *excluded* from the benefits of the contribution. This experimental design allows us to test out-group discrimination, without any confounding effects.[6]

To the best of our knowledge, our experimental protocol does not prime the ethno-linguistic identity or any other group characteristic, both in the recruitment and in the experiment itself. Subjects were recruited in a fully bilingual way, closely following the routines used in each laboratory (in Bilbao and Valencia) and university for all academic activities, and they simply chose a language for their participation. Thus, subjects naturally segregated themselves to one linguistic group or another, without any explicit or implicit reference to national identity or diversity. Subjects were not explicitly reminded of their language during the experimental sessions, and during the neutrally framed experiment, no single reference was ever made to language playing any central role in the experiment.

Experiment 1 was run in two bilingual locations in Spain: Bilbao (speaking Basque and Spanish) and Valencia (speaking Catalan and Spanish). In both locations, we manipulated the language spoken by both groups, so in some sessions the language spoken by all participants was the same (homogeneous condition) or different (mixed condition). As Section 3 describes in detail, participants were assigned to one of two rooms depending on their choice of language and were aware of the language spoken

[6] Buchan et al. (2009, 2011) were probably the first to study different nested public goods games in the field. In their game, contributing to the global public good game maximizes social welfare, but local contributions generated more returns to the own group.

in the other room through a short welcome speech made in one or the two languages in a very casual way.[7] By running the experimental sessions in two different languages and letting the subjects choose the language, we elicit their *natural* identity. As subjects were randomly assigned to the different treatments (Homogeneous or Mixed), we focus our attention on the consequences of ethno-linguistic diversity characteristics.[8] As the other (local) group may or may not share the same linguistic identity (as per their mother tongue language), we believe our game allows us to characterize the willingness to exclude the other (local) group from the benefits of cooperation depending on the ethnolinguistic identity of the other group (in homogeneous and mixed treatments) across these two bilingual societies. We collected evidence to study participants' behavioral patterns in detail, including the strength of conditional cooperation.

The results of Experiment 1 strongly suggest that ethnolinguistic diversity generates very different effects on economic performance, as diverse groups outperform homogeneous groups in one location (Valencia), and underperform homogeneous groups in another (Bilbao). These results are consistent with the existence of a strong political conflict in Bilbao (in the Basque Country most residents feel only Basque or Spanish) and the lack of it in Valencia (where most inhabitants feel members of both communities). In other words, our results are consistent with the idea that the effects of ethno-linguistic diversity are determined by a pre-existent group conflict. In the presence of group conflict (as in the Basque Country), diversity harms performance by reducing contributions to the global public good, and earnings. Conditional cooperation completely vanished in the Mixed sessions in the Basque Country, when languages spoken in both rooms were different. The opposite happened in Valencia, with strong conditional cooperation patterns in the Mixed sessions, and global contributions above those in Homogeneous sessions.

Our paper combines the results of Experiment 1 with another experiment run several months later with different participants recruited in the same locations. Experiment 2 sheds light into how empirical and injunctive norms explain the results obtained in

[7] We tried to avoid saliency of identity. Asking about the language may prime ethnolinguistic identity (see Shih et al., 1999) and for that reason we used an indirect way to select subjects in each group. This procedure should have made saliency of identity weaker.

[8] Social Identity Theory makes a conceptual distinction between language and ethnicity –see for instance Laitin (1977); Lee and Pérez (2014); Pérez (2016, 2015); and Pérez and Tavits (2017), but in this paper we observe them together and refer to them as ethno-linguistic identity, see Hogg et al. (2017) for a cautious distinction between different types of identities.

Experiment 1. Experiment 2 studies the mediating effect of *norms* on (in-group) favoritism and (out-group) discrimination, and how identity changes the empirical (or positive) *beliefs* about the behavior of others in homogeneous and mixed settings. We elicited empirical and normative beliefs in Experiment 2 with new participants with no experience on the previous study. As *observers* of the first study, they received the instructions of the first study but not the actual decisions made by the *participants*.

We elicited first-order beliefs (about the decisions made by the *participants* in the different settings) and second-order beliefs (about the decisions of the other *observers* when asked to predict the decisions of the *participants*). By focusing on the second-order beliefs, we learn about *shared beliefs*, identify differences in empirical expectations across the different ethno-linguistic groups and locations, and study normative out-group discrimination.

The elicitation of shared beliefs suggests that the positive effect of diversity in Valencia operates through optimistic (empirical) beliefs about the behavior of the out-group in Mixed sessions, while in Bilbao observers exhibited normative beliefs consistent with normative out-group discrimination. When objective information about the decisions made in Experiment I was disclosed to observers, a modest decrease in normative tolerance to discrimination is observed in Valencia (following positive news about diversity performance), while an increase in the moral justification of out-group discrimination is observed in Bilbao (following negative news about diversity performance, similar in magnitude to the positive ones disclosed in Valencia).

We also document how participants in Experiment 2 in Bilbao and Valencia react to objective information from Experiment 1 about the negative and positive effects of diversity in contribution to the global public good. In the second stage of Experiment 2, we disclosed partial information (from the first half of the first experiment) about the decisions made by the in- and the out-group, and see how observers from different ethno-linguistic groups adjust their empirical and normative expectations about the behavior of other participants. We believe it is the first experiment in which norms shifts are studied in detail. Our results are consistent with the results of Experiment 1 and 2 described above. While a substantial increase in their empirical expectations of out-group discrimination (i.e. they anticipate contribution to the global group will be lower in mixed groups) and their normative justification of out-group discrimination is

observed in Bilbao, the empirical and normative expectations of out-group discrimination is very modest in Valencia when participants in Experiment 2 receive similar, in magnitude, positive news about mixed groups.

The rest of the paper is organized as follows. In Section 2, we briefly provide background information on the field. Section 3 explains Experiment I in detail, while Section 4 presents its results. Section 5 describes Experiment II and the belief elicitation results. Section 6 concludes.

2. The bilingual context

The experiment was run in two bilingual locations in Spain: Bilbao in the Basque Country, and Valencia in the Valencian Community. In contrast with other multilingual European societies, such as Belgium or Switzerland, there is no geographical group segregation in either Bilbao or Valencia (although in rural areas Basque or Catalan is more commonly spoken). The school system is strongly based on the idea of integration, students have to study the two languages and there is contact between students from the two ethno-linguistic backgrounds (see Alexander and Christia, 2011).[9]

Table A.1 in the Appendix provides data on the percentage of the general population who speak Basque and Spanish in Bilbao, and Catalan and Spanish in Valencia, and it includes information about the two universities where this study was run. While Basque is not an Indo European language, Catalan and Spanish are, and they are much closer than Basque and Spanish.[10] Linguistic diversity has been considered a proxy for the broader notion of ethno-linguistic or cultural diversity since measuring dissimilarity between languages is relatively easy (Desmet et al. 2009).

National identity is much stronger among the Basques. According to the CIS survey (Center for Sociological Research, 2007), 48.5% of the population in the Basque Country consider themselves Basque but not Spaniards or more Basque than Spaniards, while in the Valencian Community the percentage is a mere 12.8%. Similar surveys run in different years show very similar figures. This non-identification of a large part of the Basque society with the Spanish national identity underpins the

[9] Table A.1 in the Appendix provides further information on socio-demographic characteristics of the field.

[10] See Desmet et al. (2009, 2011, 2012) for measures of distance between languages.

success of the nationalist parties in the Basque Country. In the regional elections of 2011, the Nationalist parties in the Basque Country obtained more than the 50% of the votes while this percentage was less than 15% in the Valencian Community (see Table A.2 in the Appendix). On the basis of the perceived cost of cultural heterogeneity, Desmet et al. (2011, p. 183-184) conclude that the European regions most likely to secede are the Basque Country and Scotland. Furthermore, the terrorist organization ETA has historically sought the independence of the Basque Country[11] and the establishment of the Basque as the unique national language.[12]

Although the Basque Country and the Valencian Community are both bilingual regions and share many socioeconomic characteristics, the two fields differ in the extent of conflict between their ethno-linguistic groups. Thus, they provide an appropriate ground to check the effect of diversity. We explicitly admit that we did not run the same experiment in bilingual locations with a gradient of group conflict, from inexistent to very high. Therefore, we will refrain from making any comparison between decisions across the two regions and rely on the comparative statics (how the same manipulation generates similar or different effects in both locations, and why).

Our experiments were run in two large public universities in Valencia and Bilbao. As Table A1 shows, they exhibit many common characteristics, including their size, faculty to student ratio, admission criteria, recruitment protocols and tuition fees. At the time when our experiments were run, the majority of college students in Spain attended public universities like the ones used in our study. Both the Basque Country University and the University of Valencia receive a very large majority of their students from Spain, with a very small proportion of international students (4% in the Basque Country and 6.5% in Valencia, according to Olivella, 2016). Spanish undergraduate students are not very mobile for a combination of regulatory and historic reasons, and most students in both universities come from the same region. The proportion of undergraduate students from other Spanish regions is very similar in both universities: 8.5% in the Basque Country and 8.1% in Valencia when the experiments were run (per Ministry of Education reports; the proportions have not changed much since then).

[11] Abadie and Gardeazabal (2003, p. 115) report data on the number of killings by ETA "on average, during the 1980's, ETA's activity resulted in 39 deaths per year; this figure fell to 16 per year during the 1990's. In the year 2000, ETA killed 23 people."
[12] Corcuera (1979).

3. Experiment I: design and procedures

Design

This paper reports on two different waves of experimental sessions. In this section, we describe in some detail the game played by participants in the first experiment, which was then the basis for the second experiment described in Section 5. In Experiment I, participants face a *nested* (*local-global*) *public goods game*, following the one used in Buchan el al. (2009). Participants decide each round how to allocate their endowment (100 Experimental Currency Units) between a Local public account (l_i), a Global public account (g_i) and a Private account.

Participants are randomly assigned to cohorts of 8, and each cohort of 8 participants has 2 groups of 4 participants. More formally, each individual *i* makes decisions within a group G^i. Each *group* G^i has its own Local account, and the Global public account receives contributions from the two groups of four participants (G^i and G^{-i}). Thus, while the Local public account is funded by, and benefits, members of a *local* group G^i, the Global public account is funded by, and benefits, both local groups. Therefore, the members of the group G^{-i} benefit from the contributions to the Global account made by members of the group G^i, and vice versa.

The following equation represents the individual payoff function in period *t*,

$$\pi_{it} = (e_{it} - l_{it} - g_{it}) + b_l \sum_{k \in G^i} l_{kt} + b_g \sum_{k \in G^i} g_{kt} + b_g \sum_{j \in G^{-i}} g_{jt} \qquad (1)$$

where *e* represents the participant's endowment at the beginning of the period, l_k the contribution to the Local account and g_k the contribution to the Global account. As we will explain in detail below, the two accounts yield the same marginal per capita return (MPCR).

The parameters of the game were chosen to simplify the game to our participants. The group size of both *local* groups is 4, the round endowment is 100 ECUs, and both marginal per capita returns are 0.5. The theory predicts no contributions to any of the two public accounts (Local or Global), assuming rationality and selfish preferences, as the return of the Private account is higher than the marginal per capita return of both public accounts. In the unique inefficient equilibrium of the stage game, participants make 100 ECUs.

However, cooperation may result in a much more efficient outcome. To maximize joint earnings, individuals should fully contribute to the Global account, because the social return of the Global public account (the MPCR multiplied by the number of participants benefiting from it: 0.5 x 8 = 4) is larger than the social return of the Local public account (0.5 x 4 = 2). In the efficient social welfare maximizing solution, participants fully contribute to the *global* account and make 400 ECUs (if they fully contribute to their *local* account, they would make 200).

Figure 1 depicts the dilemma faced by a participant in the experiment. We plot the social return of one unit invested in the global public good for both groups (as a dashed line) and for the local group (keeping the MPCR of the local public good fixed in 0.500). Figure 1 shows how the decision may be extremely sensitive to the MPCR chosen for the global public good (horizontal axis) if the participant exhibits some preferences for the joint social welfare of both groups. For convenience, the social return generated by a Selfish participant (investing in the private account, with normalized returns for the selfish individual and her group of 1) and a Parochial participant (fully investing in the local public good, with a MPCR of 0.5, and a social return for her local group of 2, with no benefits for the other local group) are presented as solid horizontal lines.

Figure 1 shows that any MPCR of the global public good below 0.125 poses no challenge for any participant considering the possibility of contributing to it if maximizing the social welfare of both groups. As the individual return (1, as presented by the Selfish line) is higher than the social return of the global public good (see the selfish line above the dashed line), even a social welfare maximizer should refrain from contributing to the global public good and keep the unit in her private account.

If the MPCR of the global public good is above 0.125 and below 0.250, a social welfare maximizer should abstain from contributing to the global public good and should contribute to the local one, instead. The logic is simple: while any contribution to the global public good benefits both groups, the social return for both is higher if the unit is invested in the local public good (as a Parochial individual would do; see the Parochial line above the dashed line).

If the MPCR of the global public good is above 0.250 and below 0.500, a social welfare maximizer could see that any contribution to the global public good generates

a higher social return (the dashed line is above the Parochial one). However, in this MPCR range, participants may be sensitive about the level of cooperation to the global public good of both groups. If they believe the other group is formed by parochial participants, their local group (solid line) is actually worse off than if they had invested in the local public good (Parochial line). As in Buchan et al. (2009, 2011), participants face a trade-off between the potential joint benefits of the global public good (if all participants invest in it) and the safe option of investing local, as a parochial participant would do.

[Figure 1 here]

In other words, if the MPCR of the global public good is above 0.250 but below 0.500, participants could refrain from contributing to the global public good not necessarily because they want to exclude the other group from the benefits of their contribution to the global public good, but because they fear that the contribution of the other group will be low, and want to secure the benefits of their own group.

The less interesting case is when the MPCR of the global public good is above 0.500, as even a parochial participant would prefer to invest in the global public good if trying to maximize the earnings (welfare) of her local group. The behaviorally interesting tipping point is when the MPCR of the global public good is exactly 0.500. Why? The social return of the global public good is not above (and not below) the return of the local public good. A parochial participant knows that a unit invested in the local public good benefits her local group as much as if she decides to invest the same unit in the global public good. True, the other group benefits from her contribution to the global public good and does not benefit if she decides to contribute it to the local public good. Given that the benefits for her own local group are *identical* if the unit is invested in the global public good or if invested in the local public good, a parochial participant will be indifferent if trying to maximize the benefits of her local group. If the parochial individual refrains from investing in the global public good, it must be for the only difference between both investments: by investing locally, a parochial participant may be willing to exclude the other group from the benefits of his investment. This situation only happens for a global MPCR value of 0.500.

Even when this paper benefits from the seminal work of Buchan et al. (2009, 2011, 2012), there are important differences between the experimental model used in this

paper and the paradigm used in their work. The first difference comes from the very different method used in both experiments. In the paper and pencil experiments run in six countries in Buchan et al. (2009, 2011), cooperation was measured using a variant of the nested public goods game used in our paper, called the Multilevel Sequential Contribution game (MSC). The MSC closely resembles a nested public-good game except that subjects do not make decisions directly affecting those in their groups, but the payoffs of others in future sessions. Each subject's endowment of 10 tokens is to be allocated between a Private account, a Local account and a Global account. As in our experiment, participants in Buchan et al. (2009 and 2011) could cooperate exclusively with a local group (with participants from their own locality), globally (with two other groups from other areas of their nation or other parts of the world), or with both, splitting their allocations between the local and global public goods.

Each token put into the Local account was doubled by the experimenter in Buchan et al. (2009 and 2011), so the MPCR of the Local public good was 0.5, as in our experiment. Each token allocated to the Global account was tripled by the experimenter, and equally shared by the 12 cohort members, so the MPCR of the Global Public Good was 0.25, and different from the marginal per-capita return of the global public good.

The structure of incentives of this nested public-goods dilemma is such that a selfish individual should refrain from investing in either the local or the global public goods and fully allocate her endowment to the private account. Interestingly, a social welfare maximizer should still fully invest her endowment in the global account as the social return of the Global public good is larger than the social return of the Local public good (3 versus 2). The difference is very relevant. As each local group consisted of 4 participants, and three groups or four participants could benefit from the global public good, Buchan et al. (2009 and 2011) study cooperation when a tension exists between local and global cooperation.

Contrary to our experiment, by contributing locally, participants in Buchan et al. (2009 and 2011) reduced the surplus generated and distributed to other local group members. In other words, excluding participants in other countries from the benefits of your global contribution came at the price of lower returns to local participants, a price their participants had to compare with the potential larger benefits of the global public good. In our experiment, the exclusion of the other group came at no cost for the local group,

as any unit invested in the local public good generated the same benefit for the members of the local group (0.50) than if invested globally (0.50). In other words, exclusion is costless for the in-group.

Procedures

Participants in Experiment I were students at the University of the Basque Country (UPV/EHU) and at the University of Valencia (UV). Both are public universities and the socioeconomic level of the students is similar. These universities are bilingual, which implies that the two languages are frequently used in all official communications and students typically choose to take courses or attend events in Basque/Catalan or Spanish. Subjects were electronically recruited through a bilingual e-mail, Spanish/Catalan or Spanish/Basque. The official university e-mails, and other invitations to all types of events are always sent in both languages. Subjects were given the chance to participate in experimental sessions in Basque or Spanish (at UPV/EHU) and Catalan or Spanish (at UV). Participants electronically chose the language of the session and once they came to the lab all the experimental procedures were fully conducted in that language. Framing was scrupulously neutral when mentioning the language used or chosen by subjects, without any political or identity reference.

Each session was conducted in two adjoining computer rooms (both in Valencia and in Bilbao). The main experimental manipulation comprised the selection of languages used in both rooms. While some of the sessions were *Homogeneous* (as the same language was spoken in the two rooms), others were *Mixed* (the two rooms used different languages). We use these terms (*Homogeneous* and *Mixed*) to refer to the different treatments run in each of the two locations.

All participants, in both the Homogeneous and the Mixed sessions, were requested to come at the same hour and at the same place. A very short welcome announcement was made to all the participants, in the unique language in the Homogeneous sessions, and in both languages in the Mixed sessions;[13] the participants were then directed to the previously assigned computer room. As the welcome speech was used to address the participants in the language originally chosen by them, they were aware of the language spoken by participants in the other computer room. Table 1 below describes

[13] This is also a standard procedure in these two bilingual universities.

our treatments and presents a summary of Experiment I. The language used is the same in both computer rooms in four treatments (*homogenous* sessions) and the language is different in two treatments (*Mixed* sessions).

[Table 1 here]

At the beginning of the experiment, participants are randomly matched in groups of four in each computer room and group composition remains constant throughout the session. Thus, people in the same *local* group always speak the same language. Then, one *local* group from one computer room was randomly matched with another *local* group in the other computer room. We refer to each pair of local groups, one from each room, as sections. *Sections* are fully independent from each other, so each section is an independent observation from a statistical point of view. Note that participants in one *section* may speak the same language, or a different language, depending on the treatment. Subjects were always aware of that.

After the instructions had been read aloud and before the experiment started, participants had to answer a quiz to check their understanding of the game. Then participants played 20 rounds of the public goods game. Subjects were allocated 100 ECUs (Experimental Currency Units) in each round. After each round, subject i was provided with information on their own contributions and payoffs (private, local, global and total), the sum of local contributions of group G^i, the sum of global contributions of group G^i, the sum of global contributions of group G^{-i}, and total contributions to the global account.

After the session, and while assistants were preparing envelopes with individual earnings, participants filled the 30-item Singelis subjective-individualism-collectivism scale test (Singelis, 1994; Singelis et al., 1995). We used this test to try to capture differences between ethno-linguistic groups in social norms concerning collective action. Collectivism is related to the pursuit of group interests that could be related to cooperation (Wagner, 1995). These differences are unrelated to the interaction with a different group and could be relevant for their willingness to cooperate either locally or globally.

A typical session of Experiment I lasted slightly less than two hours and the average individual payoff was around €15 (around €17 in Bilbao and €12.5 in Valencia).

4. Experiment I: Results

This section starts with some descriptive statistics and then formally tests for treatment effects and cooperation dynamics in the different treatments.

[Table 2 here]

Table 2 and Figure A.1 in the appendix show how the manipulation of diversity affects contributions to the global public good, while it has very little effect on local public good provision. In both Bilbao and Valencia, contributions to the local public good are remarkably similar across the three conditions. In Bilbao, local contributions range from 18.27 in the Mixed sessions to 20.68 in the Homogeneous sessions run in Basque (with slightly smaller local contributions in the Homogeneous sessions run in Spanish, 19.46).[14] In Valencia, local contributions range from 11.94 in the Homogeneous sessions run in Catalan, to 12.13 in the Mixed sessions and 13.96 in the Homogeneous sessions run in Spanish. These differences are never significant when running standard non-parametric tests.[15]

While there is almost no difference in local contributions in Homogeneous and Mixed sessions (differences are always lower than 10%), things are different for Global contributions. Diversity has a profound effect on contributions to the global public good. In Bilbao, global contributions in Mixed sessions are below global contributions in Homogeneous sessions (16.79 vs. 24.65 and 21.16 for Spanish and Basque, respectively). Interestingly, the effect of diversity is just the opposite in Valencia: average global contribution in the Mixed sessions is 20.67 vs. 16.37 and 13.42 in the Homogeneous sessions (for Spanish and Catalan, respectively).

When there is ethno-linguistic diversity, as implemented in Experiment I, global contributions are reduced 27% in Bilbao, and increase 38% in Valencia. While the substantial difference in global contributions in Valencia is not statistically significant using a Mann-Whitney non-parametric test with one observation per each of the 6

[14] Local contributions in Bilbao are not significantly different in the Mixed and Homogeneous sessions, using a Mann-Whitney non-parametric test with one observation per section. The difference between local contributions in the Homogeneous sessions run in Basque and in Spanish is also not significant (z= 1.086; p-value=0.2774).

[15] The local contributions of Catalan and Spanish participants in Homogeneous sessions are not statistically different in Valencia: z= -0.320; p-value=0.7488.

sections (Homogeneous vs. Mixed: z =-1.218; p-value=0.2234), diversity significantly reduces global contributions in the Basque Country (z= 1.903; p-value=0.0570).[16]

Figure 2a shows how global contributions evolve over the 20 periods of the experiment by diversity levels and location. While contributions are initially very close in Valencia, and then the Mixed sessions outperform the Homogeneous in Valencia, in Bilbao the difference starts from period 1, and continues until the end of the experiment. The decline in contributions is higher in Valencia than in Bilbao, as we will confirm in the econometric analysis below.[17] Figures 2b and 2c presents contributions to the local public good and the private account, by location and treatment (homogeneous and mixed), mirroring the results of Figure 2a: contribution to the private account is remarkably above (below) in mixed sessions in Bilbao (Valencia) than in homogeneous ones, while no clear pattern emerges from contributions to the local public good (analyzed below)

[Figure 2a here]

[Figure 2b here]

[Figure 2c here]

Table 3 confirms these effects on the contribution to the Global public good game by estimating five linear regression models. Models (1) and (3) show the effect that the Mixed condition has in each of the two locations, while Model (5) checks for significant differences between both locations. Models (2) and (4) introduce the language condition in the analysis to control for differences between the two linguistic groups concerning the effect of the Mixed treatment. As contributions to the global public good may be related to the importance of collective vs. individual action in each culture, we control in all the models presented in Table 3 for any differences between ethno-linguistic groups in collectivism and individualism (Singelis, 1995; see the appendix for a detailed description).

---

[16] It should be noted that the average global contribution is not significantly different for the Homogeneous sessions run in Spanish and in Basque (z= -0.575, p-value=0.5653) and in Spanish and Catalan in Valencia (z=-0.961, p-value=0.3367)

[17] Figure A2 in the appendix shows how language makes a difference in the Mixed sessions run in Bilbao and Valencia. Even though culture differences between the different ethno-linguistic groups is not the objective of this paper, it is interesting to note the existence of some differences: while in the BC the Basque speaking participants contribute similarly in the Homogeneous and the Mixed sessions (remarkably so at the first half of the 20-round game), Spanish speaking participants contribute less, and from the very first period, in the Mixed sessions, not reacting to the initially higher contributions of the other group.

The results presented in Table 3 for the panel data regressions[18] show that diversity does not necessarily reduce contributions to the global public good, and that the effect may be different depending on the linguistic group. In Valencia, the coefficient of *Mixed* is not significant in Model (1), but controlling for *Language* in Model (2) we find that the difference between the Mixed and the Homogeneous sessions (see the coefficient of *Mixed + Mixed*Language* at the lower part of the table) is positive and significant, meaning that we find higher contributions in the Mixed sessions for the Catalan group. This is not the case for the Spanish group as the coefficient of *Mixed* in Model (2) is not significant.

In Bilbao, the coefficient of *Mixed* is significant and negative in Model (3), suggesting that in Bilbao ethno-linguistic diversity leads to a lower level of cooperation. Controlling for *Language* in Model (4) allows us to see any possible differential behavior between the linguistic groups. We find that the difference between the Mixed and the Homogeneous sessions (see the coefficient of *Mixed + Mixed*Language* at the lower part of the table) is not significant, implying that behavior in the Mixed and Homogeneous sessions for the Basque group are very similar. However, *Mixed* is significant in Model (4) meaning that the effect of diversity is strong and negative for the Spanish group, driving the total negative effect observed in Model (3).

The differences between the two locations concerning the effect of diversity are statistically significant, as indicated by the interaction term of *Mixed+Mixed*BC* in Model (5). On average, the contribution to the global public good game in the Mixed sessions in Bilbao is around 6.8 ECUs lower than in the Homogeneous sessions. The effect in Valencia is measured by the coefficient of *Mixed* in Model (5), and we find that the difference between the Mixed and the Homogeneous sessions is not significant.

It is worth noting that the higher level of contributions observed in Bilbao, as compared to Valencia, in the Homogeneous sessions (almost 10 ECUs higher, see Model (5)),[19] vanishes when Basque and Spanish interact. Diversity in the BC breaks the high level of contributions. The coefficient of *BC+Mixed*BC* is not significant

[18] We also study the possible treatment effects using Tobit and Fractional Multinomial Logit. See Table A3 in the appendix. We found the same results as in Table 3.

[19] We do not focus on this difference in the paper, as our main point is the difference between Mixed and Homogeneous settings. However, this result is in line with previous literature; see Criado et al. (2015, 2018).

indicating that the levels of contributions in the two locations in the Mixed sessions are not different.

[Table 3a here]

Note that we cannot exclude the possibility of some group identity emerging in the Homogeneous sessions. In our experimental protocol, participants are always taken to one of two rooms, assigned to a group of four people in the same physical space, different from the other group in their section. Homogeneous sessions may trigger some group effect even when the framing was neutral. As these differences do not seem to be driven by cultural differences across the different linguistic groups, we simply consider the Homogeneous sessions as a baseline benchmark. We measure any *additional* group effects in the Mixed sessions (that is, above the potential group effects in the Homogeneous sessions), with the introduction of the dummy *Mixed* and its interactions in Table 3.

[Table 3b here]

Table 3b confirms these effects on the contribution to the Global public good game by estimating six linear regression models. Models (1) and (3) show the effect that the Mixed condition has in each of the two locations. Models (2), (3), (4) and (5) replicate the same analysis than model (1) and (2) by participant's language. We again introduce a control variable (*Selfish*) that measures the degree of selfishness of each participant and a new variable (*Private*) that control for the amount participants kept in the private account. The two variables are also introduced as interaction term. As contributions to the global public good may be related to the importance of collective vs. individual action in each culture, we control in all the models presented in Table 3 for any differences between ethno-linguistic groups in collectivism and individualism (Singelis, 1995; see the appendix for a detailed description).

The results presented in Table 3b for the panel data regressions[20] show that diversity does not necessarily reduce contributions to the global public good, and that the effect may be different depending on the linguistic group. In Valencia, the coefficient of

[20] We also study the possible treatment effects using Tobit and Fractional Multinomial Logit. See Table A3 in the appendix. We found roughly the same results as in Table 3.

Mixed is not significant in each of the three models. We find that the difference between the Mixed and the Homogeneous sessions (see the coefficient of *Mixed* + *Selfish*Mixed* + *Private*Mixed* at the lower part of the table) is positive although not significant. We also find the same result for the Spanish and Catalan group as indicated in model (2) and (3).[21]

In Bilbao, the coefficient of *Mixed* is negative in Model (4), suggesting that in Bilbao ethno-linguistic diversity leads to a lower level of cooperation. However, the behavior between linguistic groups (Spanish and Basque) is different. We find that in model (5) the difference between the Mixed and the Homogeneous sessions (see the coefficient of *Mixed* + *Selfish*Mixed* + *Private*Mixed* at the lower part of the table) is negative and significant, implying there is a treatment effect for the Spanish group, meaning that the effect of diversity is strong and negative for those participants. However, this effect is not statistically significant for the Basque group.

[Table 3c here]

Table 3c confirms these findings studying the contribution to the local public good as a function of our manipulation (homogeneous vs. mixed), the location and language. Not surprisingly, the same effects observed in Tables 3a and 3b for contributions to the global public good (negative and substantial effects of diversity in Bilbao, positive and insignificant effect of diversity in Valencia) are reversed when exploring the contributions to the local public good.

Table 4 analyzes efficiency using earnings as a proxy. We study whether the earnings of participants in the different treatments significantly differ, as there are different group externalities and efficiency gains to exploit in this game. It could be, for instance, that any efficiency losses generated by lower contributions to the global public good were compensated by higher contributions to the local public good. Earnings are the outcome of cooperation and we are interested in the performance of Mixed vs. Homogeneous societies, as discussed in the introduction. If diversity harms economic performance, we could align the result with the idea that Homogeneous or well integrated societies are more efficient because diversity harms cooperation. Given

our experimental method, we can also learn whether both groups benefit from cooperation, and which one benefits more.

Table 4 presents the same panel data analysis as before but using earnings as the dependent variable. We observe from Model (1) that in Valencia there is no significant change in average earnings when people interact with the same group or with people from a different linguistic community. In line with previous results, Model (2) shows a weakly significant increase in earnings in the Mixed sessions (the coefficient of *Mixed+Mixed*Language* in Model (2) is significant at 10%) for Catalan participants.

The analysis of performance in Bilbao is quite different. On average, participants in the Basque Country earned considerably less in the Mixed sessions, when Spanish and Basque interact, than in the Homogeneous condition (Model (3)). Nevertheless, in Model (4) we see that the Basque linguistic group's earnings are not significantly lower in the Mixed sessions (coefficient of *Mixed+Mixed*Language*) meaning that the aggregate decrease in earnings comes from the negative effect of diversity on the Spanish linguistic group.

Model (5) suggests that participants in Bilbao get significantly higher payoffs than in Valencia (coefficient of *BC*). However, the efficiency losses in the Mixed sessions in Bilbao are large and significant, as the coefficient of *Mixed+Mixed*BC* indicates. Efficiency gains obtained in the Homogeneous sessions in Bilbao (relative to Homogeneous sessions in Valencia) are completely wiped out by the group conflict (*BC+Mixed*BC* is not statistically significant).[22]

[Table 4 here]

We investigate now the dynamics of conditional global contribution to shed some light on the main behavioral factors driving these treatment differences. We use a linear regression model, where the dependent variable is the subject's global contribution in period *t*. The explanatory variables are: *Period*, to capture the possible trend in contributions; $Contribution_{i,t-1}$, the individual global contribution in the previous period, to track any possible persistence; $Contribution_{g,t-1}$, the out-group average global

[22] In the homogenous sessions in the BC participants obtained around 20% higher payoffs than in the VC.

contribution in the previous period; and $Contribution_{I\text{-}i,t\text{-}1}$, the average global contribution in the previous period of the members of the in-group excluding *i*. Figure 3 presents the linear prediction of the individual global contribution in each period ($Contribution_{i,t}$) on the previous contribution of the out-group to the global public good ($Contribution_{g,t\text{-}1}$). Table 5 presents the regression results controlling also for the $Contribution_{I\text{-}i,t\text{-}1}$.

[Figure 3 here]

[Table 5 here]

Figure 3 and Table 5 show a dramatic difference in conditional cooperation patterns between the two locations. As expected, there is persistence in global contributions, meaning that $Contribution_{i,t\text{-}1}$ is positive and highly significant in all treatments. This persistence does not apply to the lagged global contribution of the in-group ($Contribution_{I\text{-}i,t\text{-}1}$ is never significant), suggesting that global contributions are driven by the behavior of the out-group rather than by the in-group. Conditional cooperation is confirmed by the positive slopes of the fitted lines in Figure 3 and the positive and significant coefficients of $Contribution_{g,t\text{-}1}$ in Table 5 with one remarkable exception: Mixed sessions in Bilbao.[23]

While the conditional cooperation dynamics in the Homogeneous session in Bilbao and in Valencia are not very different (Figure 3: SS, BB, CC), diversity again plays very different roles in both locations. The dotted line corresponding to Mixed sessions has a larger slope in Valencia (SC), and it is remarkably flat in Bilbao (SB), suggesting that participants in Bilbao do not react to the contribution of the out-group to the global public good. Table 5 confirms this, as the coefficient of the $Contribution_{g,t\text{-}1}$ variable is significantly different from 0, with the exception of the Mixed session in Bilbao. In the Mixed sessions in Bilbao, participants do not react to the lagged global contribution of the out-group, nor do they react to the lagged global contribution of their in-group, and

[23] A similar effect appeared in Keuschnigg and Schikora (2014), where conditional cooperation (defined as the effect of a leader's contribution) was absent in heterogeneous groups.

global contribution does not decline (another exception to the general rule, as the coefficient of *Period* is significant and negative in all the other cases). A sharp contrast with the large and significant coefficient of $Contribution_{g,t-1}$ in Valencia (the largest of all four estimations), in line with previous results.

This result points to the existence of different rules of behavior that preempt conditional cooperation in Bilbao. Such rules could explain why participants do not react to the observed out-group behavior, and why decisions do not change over time. In the next section we explore this issue, introducing Experiment II.

5. Experiment II: Empirical and normative (shared) beliefs

The results obtained in the first experiment show a significant and detrimental effect of diversity on contributions and efficiency in Bilbao, and a positive effect in Valencia (a marginally significant positive effect among Catalans, and a strongly significant conditional cooperation pattern). Both results deserve a deeper analysis. The effect of diversity under group conflict is roughly consistent with the literature on in-group favoritism, and out-group discrimination. Beyond the different effect of diversity on contributions to the global public good, the extremely weak conditional cooperation pattern found in Mixed sessions in Bilbao suggests that participants in Experiment I do not respond to the outgroup behavior but to some exogenous norm.

In Experiment II we explore the role of social norms our participants may have brought with them to the laboratory with new experimental sessions run in the same locations with different subjects. Third-party beliefs elicitation avoids some problems identified in the literature when behavior and beliefs come from the same participants[24] as well as any potential hedging problem.[25] Our objective is to elicit behavioral norms at two different levels (empirical: subjects' beliefs on how people behave, and normative: subjects' beliefs on how people should behave) using shared second order beliefs as a proxy. Empirical and normative norms are also known as descriptive and injunctive norms (see Keizer et al., 2008).

[24] One of the problems is that any belief elicitation previous to the experiment may change participant behavior, although the literature is not clear about the direction of the effect. For instance, Croson (2000) found in a public good experiment that eliciting beliefs leads participants to decrease their contribution, while Gaechter and Renner (2010) found the opposite. See also Brañas-Garza and Espinosa (2011).

[25] When participants are paid for their actions and their beliefs, risk-averse subjects may want to use their stated beliefs to hedge against bad outcomes of their decisions in the game (see Blanco et al., 2010).

Participants in Experiment II will be making decisions under two different information sets. First, without having any information about the decisions of participants in Experiment I. Then, without being aware in advance that this information would be disclosed, participants receive reliable and objective information about the behavior of participants in the first ten rounds of Experiment I, and their empirical and normative expectations were measured again. So, in addition to the direct measurement of empirical and normative expectations, we also study how norms adjust to objective information about the behavior of individuals in Experiment I, to verify if norms are adjusted. We explain the experimental design and procedures in subsection 5.1, the results of the first block of decisions (with no information) in subsection 5.2, and we explore the evolution of norms, once the objective information is disclosed, in subsection 5.3.

5.1 Experimental design

The experimental procedures in Experiment II followed closely those in Experiment I, and follow the paradigm of Bicchieri (2006), as we will be eliciting empirical (descriptive) and normative (injunctive) expectations about the decisions of other participants in Experiment I and II. The difference between both expectations is simple: while empirical expectations refer to the expected behavior of other individuals in the same situation (we expect others to wait in line, or to contribute to the global public good), a normative expectation incorporates a moral judgement about the appropriateness of certain behavior (we expect others to believe I ought to wait patiently, or contribute myself). Bicchieri (2006) argues that social norms are behavioral rules supported by a combination of empirical and normative expectations. Our second experiment specifically aims to capture whether exclusion patterns in our nested public goods game are associated to an empirical expectation, a normative one, or both.

Experiment II was run some months after Experiment I, in 7 additional sessions with 140 new participants. A typical session in Experiment II lasted less than one hour and the average payoff was €12.65 in the BC and €13.29 in the VC. Participants were recruited through the same bilingual e-mail used in Experiment I, and they were asked to choose a session run in one of the two languages (Spanish or Basque/Catalan). The

subjects who participated in Experiment II had not participated in Experiment I. Once subjects arrived at the laboratory, they received the same instructions as in Experiment I, and had to complete a very similar quiz to make sure they understood the game. The main difference was that this time they did not actually play the public good game, but rather they had to guess the decisions made by participants in Experiment I. For the sake of clarity, we will call subjects in Experiment I *participants*, and subjects in Experiment II *observers*.

The *observers* were asked to make four different types of decisions, first without information about the decisions made by participants in Experiment I, and then with some information, as we will explain later in this section. We describe in some detail the four sets of decisions:

FIRST SET OF QUESTIONS. *First order empirical beliefs:* in each location (Bilbao and Valencia), and once the game played by *participants* in Experiment I was carefully introduced to them, *observers* were asked to predict the average contributions to the local and global public goods in the different conditions described in Table 2. That is, they have to predict the decisions made in the Homogeneous sessions run in Spanish (Decision 1), the Homogeneous sessions run in Basque/Catalan (Decision 2), and the decisions made in the Mixed sessions by Spanish speaking participants (Decision 3) and Basque/Catalan speaking participants, depending on their location (Decision 4).[26] All these decisions were incentivized, being the reward proportional to the deviation between their individual prediction and its true value.[27]

SECOND SET OF QUESTIONS. *First order normative beliefs:* non-incentivized decisions with a normative component. *Observers* were asked to reveal what they believe *participants* should have done in the 4 decisions described above. Answers could not be incentivized because of their normative component that is linked to a purely moral judgement.

Given that our objective is to map the beliefs *shared* by the different linguistic groups in the different locations, in the third and fourth sets of questions *observers* were asked to predict the decisions of other *observers* in the first two sets of questions, using the same incentives described above.

[26] Note that *observers* in Experiment II were never asked to make a prediction about the decisions made in the other location. In the BC, *observers* were asked about decisions in the BC, and *observers* in the VC were asked exclusively about decisions made in the VC.
[27] See the experimental protocol in the Appendix; observers were also paid a show-up fee of €5.

THIRD SET OF QUESTIONS. *Second order empirical beliefs: observers* predicted the decisions made by other *observers* in the first set of questions.

FOURTH SET OF QUESTIONS. *Second order normative beliefs: observers* predicted the decisions made by other *observers* in the second set of questions.

Table 6 summarizes the four sets of decisions made by a representative *observer*. Experiment II maps shared expectations (2nd order beliefs), in each location, for each condition run in Experiment I, at two different levels (*empirical and normative*). We will focus on the analysis of 2nd order beliefs, as proxies for the descriptive and injunctive norms (shared *empirical* and shared *normative* expectations, following Bicchieri, 2006).

[Table 6 here]

5.2 Experimental results

We present the results of Experiment II in Table 7, considering only the beliefs about contributions to the global public good. Our first task consists in analyzing whether the general results of Experiment I and II are mutually consistent.

[Table 7 here]

Table 7 presents the means and standard deviations of the empirical and normative 2nd order beliefs obtained in Experiment II, with regards to contributions to the global public good. In the first column (*Own*), we present the beliefs of (Spanish, Catalan, Basque) observers when interacting with their in-group in Homogeneous sessions. We consider this natural reference point the *benchmark* for each location, and linguistic group. In the other columns, we present the empirical and normative expectations of (Spanish, Catalan, Basque) observers. The second column (IG Mixed) shows the beliefs about the in-group when interacting with the other group in the Mixed sessions; column 3 (OG Mixed) presents the beliefs about the out-group in the Mixed sessions.

We hypothesize that participants may hold different shared *empirical* or *normative expectations* (as captured by their 2nd order beliefs) about the behavior of their in-group in Homogeneous session (when matched with other members of their in-group) and in Mixed session (when matched with members of the out-group), as observed in

Experiment I. Following a similar logic, we hypothesize that the behavioral patterns observed in Experiment I could be due to different expectations about the out-group, empirically or normatively.

If our hypothesis held true, beliefs would at least partially reproduce, with different participants and at a different time, the treatment differences observed in Experiment I: While diversity had a negative effect on global contributions and efficiency in Experiment I in Bilbao, it had a weak positive effect on global cooperation in Valencia. The positive (negative) effect of diversity in Valencia (Bilbao) could be due to optimistic (pessimistic) empirical expectations about the behavior of the in-group, and/or the out-group in Mixed sessions (relative to their behavior in Homogeneous sessions), or to the existence of a positive (negative) normative discrimination.

Our first analysis consists in assessing whether the language spoken by observers made a significant difference in their estimation of empirical and normative expectations of their in-group and the out-group, with significance denoted by crosses (+) in Table 7. Interestingly, both linguistic groups make extremely consistent predictions across all conditions in Valencia, as the lack of crosses (+) between rows suggests. In other words, the expectations of Spanish and Catalan observers in Valencia are never significantly different.

However, the expectations of Spanish and Basque observers in Bilbao differ in what they predict for their respective outgroups in Mixed sessions: empirically, Spanish observers have a much more optimistic expectation (38.47) about their out-group (Basque individuals in Mixed sessions) than Basque observers have (27.22) about their out-group (Spanish individuals in Mixed sessions). In a similar fashion, the normative expectations of Spanish observers about their out-group (Basque subjects) is much higher than the normative expectations of Basque observers: while Spanish observers believe the out-group should contribute 45.16 to the global public good, Basque observers think Spanish subjects should contribute substantially less (30.27, although the difference is only marginally significant).

These results indicate that, while diversity is not associated to discrepancies between Spanish and Catalan observers in Valencia, there are significant differences between Spanish and Basque observers when forming beliefs about the out-group in Mixed sessions. Moreover, the beliefs of Basque observers about the out-group are consistent

with what actually happened in Experiment I (Spanish participants contribute less than their out-group in Mixed sessions).

Another useful way of analyzing the data from Experiment II is to consider differences between expectations on behavior of the in-group in Homogeneous sessions, and the behavior of the in-group and the out-group in Mixed sessions. In other words, we compare what an *observer* considers to be the (empirical or normative) belief for her in-group, when matched with other in-group members, with what the same observer considers to be the (empirical or normative) belief in Mixed sessions, for their in-group and the out-group. Table 7 presents the results of non-parametric tests (significance is denoted with asterisks *), performing these comparisons.

The results yield some light about the nature of the diversity effects in Valencia and Bilbao. In Experiment II we observe positive and significant differences between the empirical expectations of observers in Valencia, consistent with the positive effect of diversity. Both Spanish and Catalan observers believe the out-group will contribute more to the global public good than their own benchmark: 41.3 versus 30.57 for Spanish observers and 38.57 versus 30.47 for Catalan observers. Diversity also has some effect on expectations about the in-group; while Spanish observers do not expect their in-group to contribute less in Mixed sessions (34.30 versus 30.57, but insignificant), Catalan observers expect their in-group to contribute more when matched with Spanish (34.30 versus 30.47, significant at the 5% level).

Differences in normative expectations in Valencia with respect to the benchmark are weakly significant for Catalans for their out-group and not significant for Spanish, which suggests that the positive effect of diversity in Valencia may be associated with optimistic expectations about global contributions in Mixed sessions, particularly for the out-group, more than with the existence of a strong positive normative discrimination.

Results are different in Bilbao. Relative to the benchmark (what Spanish and Basque observers expect of their in-group in Homogeneous sessions), neither Basque nor Spanish observers form statistically different empirical expectations about their in-

group or the out-group in Mixed sessions. In other words, observers believe individuals will contribute the same in Homogeneous and Mixed sessions. [28]

Poor performance in Mixed sessions in Bilbao seems to be associated with normative discrimination. Table 7 documents that Basque observers' beliefs show normative discrimination by the out-group. The pattern is interesting for two reasons. First, Spanish observers believe subjects should not contribute less than their benchmark (40.29) in Mixed sessions (43.42 if Spanish, 45.16 if Basque; no difference is significant). Second, relative to their benchmark (40.27), Basque observers believe Spanish participants *should* contribute less in Mixed sessions, 30.27 vs. 40.27. We could interpret this as an example of tolerance towards exclusion from the out-group: Basque observers normatively accept being discriminated by the out-group.

Note that the normative expectations by Basque observers are coherent with the actual outcome shown in Table 3: in the Mixed sessions Spanish participants contributed less than Basque participants in Homogeneous sessions.

Figure A.3 in the appendix presents the results of Experiment II graphically, at the empirical and normative level, respectively, using kernel densities. While empirical beliefs positively react to diversity in Valencia (as suggested by the shift to the right in the distribution when Mixed), there is no such a shift in Bilbao (if any, a weak shift to the left among Basque, when Mixed). Consistently with Table 7, the bottom half of Figure A.3 shows no significant shift in normative beliefs in Valencia, and a shift to the left in Bilbao (only among Basque observers), consistent with a shared belief on the appropriateness of Spanish contributing less when matched with their out-group.

To summarize this analysis, the results of Experiment II in Valencia point to optimistic empirical expectations about the out-group behavior in Mixed sessions (compared to Homogeneous sessions) as a factor driving the positive results of diversity, while in the other location we do not find such optimistic empirical expectations and indeed diversity did not have a positive effect. In Bilbao, the different expectations of Basque observers about the behavior of the in-group and the out-group in the Mixed sessions (they expect a lower contribution by the out-group), as well as the normative beliefs

[28] Here the benchmark is the contribution in the Homogeneous sessions. However, note that Basque observers expect Spanish *will* contribute less than the Basque in Mixed sessions (27.22 versus 34.09, significant at 5%).

indicating that the out-group should contribute less (compared to the in-group behavior in Homogeneous sessions) are consistent with the actual outcome of Experiment I and the negative effect of diversity.

5.3 Norm adjustment

As explained in the introduction to Section 5, *observers* in Experiment II received information about the average decisions made by each group of individuals (Spanish and Catalans in Valencia, Spanish and Basque in Bilbao) for each condition (Homogeneous and Mixed sessions) and decision (contribution to the local and global public good) in the first ten rounds of Experiment I. Following the same protocol described in section 5.1, observers were asked to answer the same four sets of decisions, now about the second half of Experiment I (rounds 11 to 20), in the same order, without receiving any feedback about their earnings, or the decisions of observers in the first part of Experiment II.[29]

Table 8 presents the results of the last part of Experiment II, by location (top half for Valencia, bottom half for Bilbao), expectations type (empirical and normative, as labelled in the first column), linguistic group of observers (Spanish and Catalan in Valencia, Spanish and Basque in Bilbao), and by the conditions of the sessions (Homogeneous Spanish, Homogeneous Catalan/Basque, Mixed). For the experimental sessions run in Valencia, we present first the 2nd order empirical and normative expectations elicited in the first part of Experiment II (as presented in Table 7), then the signal received by observers, and finally the 2nd order empirical and normative expectations elicited *after* receiving information about the decisions made by participants in the first ten rounds of Experiment I.[30] Table 8 includes information about contribution to the global public good, pooling both linguistic groups in Mixed sessions.

[Table 8 here]

The information received by *observers* about Experiment I followed two clear patterns. First, it reflects the positive effect of diversity in Valencia (contribution in the Mixed

[29] As explained, observers were aware that they would be asked to make additional decisions once the first four sets of questions were completed. However, they did not receive information about the content of the last part of the experiment.

[30] In Experiment II, observers received information about local and global contributions, disaggregated by linguistic groups in the Mixed sessions.

sessions, 22.59, above the contribution in Homogeneous sessions, 18.97 and 17.06), and the negative effect of diversity on contributions to the global public good in Bilbao (18.09 in Mixed versus 26.15 and 22.41 in Homogeneous sessions). Second, the actual contributions of *participants* in Experiment I are always substantially below the empirical expectations of *observers* in Experiment II.

The first result comes from the comparison between the 2nd order empirical and normative expectations of observers *before* and *after* the information is released, as presented in the adjustment rows. While the difference between empirical and normative expectations, before and after receiving the information is significant in Valencia (see Table 8), the pattern is less consistent in Bilbao; for Spanish observers adjustments are significant with the exception of normative beliefs in the Spanish Homogeneous sessions; for the Basque observers, adjustments tend to be less significant. Indeed, Basque observers react very little to the distance between their original (2nd order) empirical or normative expectations and the actual behavior of participants in Homogeneous sessions. Basque observers seem to give much more weight to prior beliefs than to the new information provided, so that there is no significant adjustment in normative beliefs. This importance of prior beliefs in the formation of expectations may be behind the lack of conditional cooperation; that is, the observation of contributions in the previous period has little weight in the formation of the posterior beliefs; thus, beliefs are more persistent over time and so are contributions (see Figure 1).

We finish the analysis of Experiment II exploring the predominance of out-group discrimination in the different groups, and how out-group discrimination survives the disclosure of objective information. Let us first specify the implications of out-group discrimination in this context. Out-group discrimination is detected in Experiment II when an observer's empirical or normative expectations about the global contribution of members of her in-group when matched with other in-groups (g^{homo}_{it}) is higher than when matched with out-groups (g^{mixed}_{it}), as expression (2) summarizes.

$$g^{homo}_{it} > g^{mixed}_{it} \qquad (2)$$

In other words, a Spanish (Catalan, Basque) observer has an empirical expectation of out-group discrimination if she believes that Spanish (Catalan, Basque) in her location *will* contribute more in Homogeneous sessions than in Mixed ones. Similarly, we

consider that a Spanish (Catalan, Basque) observer has a normative expectation of out-group discrimination if she believes Spanish (Catalan, Basque) in her location *should* contribute more in Homogeneous sessions than in Mixed ones. We interpret this normative expectation as moral tolerance for out-group discrimination.

Experiment II does not only allow to analyze the prevalence of (empirical and normative) out-group discrimination in each location and group of observers, but also how this proportion changes when receiving objective information about the behavior of participants in Experiment I. Figure 3 plots the percentage of observers expecting out-group discrimination as defined above, for both locations (Valencia and Bilbao), pooling the 2nd order expectations of observers in two polar cases: when forming expectations about their own group (Spanish expectations about Spanish; Catalan expectations about Catalan; Basque expectations about Basque), or about the other group (e.g. Spanish about Catalans in Valencia and about Basque in Bilbao). Finally, Figure 4 presents the percentage of observers with discriminatory expectations before (O) and after (1) the disclosure of information.[31]

[Figure 4 here]

[Table 9 here]

Figure 4 can be analyzed with the help of Table 9, which includes some basic non-parametric tests. First, before the information was released, the expectations of out-group normative discrimination (measured as percentage of observers) are quite similar in Valencia and Bilbao, between 35% and 30% of discrimination. Expected empirical discrimination in the own group is also similar (30% and 36.67%), but the expected empirical discrimination in Bilbao for the out-group is 46.67%, significantly above the 31.25% in Valencia. In other words, with no information about participants, in Valencia and in Bilbao observers exhibit very similar levels of normative discriminatory expectations in their own group and the other (differences are never significant, as Table 9 shows), and the same is true for empirical discrimination in the own group. In sum, while the averages of expected empirical and normative out-group discrimination (by the in-group and the out-group) before any information is disclosed is around 30%-36.67%, in Bilbao the empirical average for the out-group behaviour is significantly larger (46%); these beliefs would be consistent with a lower level of

[31] Figure A.4 in the appendix plots every possible combination of observers and participants' linguistic groups with very similar results.

contributions in the Mixed sessions (compared to the Homogeneous sessions), as it was the case in Experiment I in Bilbao.

As Table 9 shows, the effect of information on the percentage of observers expecting discrimination is always significant (see the top half of the table, when comparing variables labelled B, before, and A, after, the information was released). However, the magnitude of the effect is not: while in Valencia the positive news generates a modest reduction of discriminatory expectations, the negative news generates a dramatic increase in discriminatory expectations (empirically) and discrimination tolerance (normatively) in Bilbao. In the later case, the proportion of observers empirically expecting out-group discrimination when considering their own group (Spanish observers predicting discrimination among Spanish, Basque observers predicting discrimination about Basque subjects) goes up to 96.67% (empirically). The proportion of observers normatively justifying discrimination goes up to 70%.[32]

In sum, information disclosure has an asymmetric effect on beliefs in Valencia and Bilbao. Consistent with the information revealed on the level of contribution of the different groups, beliefs in Valencia adjusted to a lower level of out-group discrimination and to a higher level in Bilbao, but the reaction in the second location was of a much larger magnitude, both for empirical and normative expectations.

6. Concluding Remarks

The relationship between ethno-linguistic diversity and economic performance is becoming more relevant in a world where economic relations involve groups of increasing diversity, and migration is in the political agenda in Europe and the US. Indexes of ethnic fractionalization are often used to empirically estimate the consequences of diversity on macroeconomic performance and growth,[33] although beyond the number of ethnic groups, the conflict between the different groups is not always well captured by these indexes (Fearon, 2003). Montalvo and Reynal-Querol (2005) propose an index of polarization as a better measure of conflict, and

[32] The results are very similar when analysing 1st order normative expectations, based on individual moral judgements: around 55% of observers in Bilbao answered that their own group should contribute more alone than Mixed with the out-group, see Figure A4 in the Appendix.

[33] Easterly and Levine (1997), La Porta et al. (1999), among others.

Gardeazabal (2011) documents the relationship between linguistic polarization and armed conflict in the Basque Country.

This paper uses a novel research method to study the impact of ethno-linguistic diversity on economic performance through its impact on cooperation. By running experiments in two bilingual locations in Spain, we carefully keep some variables under control and study how randomly assigned participants react to a precise manipulation of ethno-linguistic diversity. As our experiment is run in two bilingual locations with a very different level of group conflict, we conjecture that this different environment could shape the effect of diversity on economic performance.[34] The results of Experiment I in Valencia strongly suggest that diversity does not necessarily diminish the provision of global public goods, and it may reinforce conditional cooperation patterns. Provision, and efficiency, is significantly reduced in Bilbao, a location with a larger group conflict.

Using their mother tongue language, we assign participants to different treatments and groups within each treatment, thus obtaining interactions between naturally occurring groups and identities. By designing two waves of experiments, we explored the rationale behind some findings of Experiment I carefully constructing two sets of empirical and normative expectations, using 2nd order beliefs.

Our results indicate that when integration is not an issue, and group conflict is at most weak, diversity may generate a self-fulfilling *empirical* prophesy: both Spanish-speaking and Catalan-speaking participants exhibit overoptimistic beliefs about the contributions of the out-group under diversity, and do contribute more to the global public good in Mixed groups than in Homogeneous settings without any out-group discrimination being observed. In Bilbao, we document how the negative effect of diversity is associated with normative discrimination compatible with a willingness to exclude the out-group from the benefits of global contribution.

Using a simple manipulation, we also circulate objective information about the actual decisions of participants in Experiment I to observers in Experiment II. Not surprisingly, empirical and normative expectations are adjusted to the real decisions of participants in Experiment I. However, while the good news about the performance of Mixed sessions in Valencia, generates a modest reduction of tolerance to

[34] Note that conflict per se is not included in the design of the experiment and therefore we only test formally the effect of diversity in the different locations.

discrimination, the bad news about the performance of the Mixed sessions in Bilbao triggers a massive increase in both expected out-group discrimination and the tolerance to it.

To conclude, we show in this study that diversity does not always have to destroy cooperation between groups, and that working to reduce group conflict may help to reduce any under-provision of global public goods. Our results suggest that more integration could mitigate or reverse the pernicious consequences of ethno-linguistic diversity on macroeconomic performance. The literature on conflict resolution has pointed out that institutions that give protection to minorities may play a role (see for example Easterly, 2001) and also that more communication between the conflicting groups may contribute to eliminate prejudice and improve cooperation (the intergroup contact hypothesis; Allport, 1954).

Acknowledgements
We thank Javier García, Yolanda Chica, Cristina Pizarro-Irizar and Alaitz Artabe for their help in running the experiments. Financial support from the Spanish Ministry of Economic Affairs and Competitiveness and Fondo Europeo de Desarrollo Regional (ECO2015-64467-R, MINECO/FEDER), the Basque Government (DEUI, IT783-13), UPV/EHU (UFI 11/46 BETS) and the ESRC Network for Integrated Behavioural Science (NIBS) is gratefully acknowledged.

References

Abadie, A. (2006). Poverty, Political Freedom, and the Roots of Terrorism. *American Economic Review* 96(2): 50-56.

Abadie, A., Gardeazabal, J. (2003). The Economic Cost of Conflict: A case study of the Basque Country. *American Economic Review* 93(1): 113-132.

Abadie, A., Gardeazabal, J. (2008). Terrorism and the World Economy. *European Economic Review* 52 (1): 1-27.

Abbink, K., Brandts, J., Herrmann, B., Orzen, H. (2010). Inter-group conflict and intra-group punishment in an experimental contest game. *American Economic Review* 100(1): 420-447.

Aguiar, F., Branas-Garza, P., Espinosa, M.P., Miller, L. (2010). Personal identity. A theoretical and experimental analysis, *Journal of Economic Methodology* 17(3): 261-275.

Ahn, T. K., Isaac, R. M., & Salmon, T. C. (2011). Rent seeking in groups. *International Journal of Industrial Organization*, *29*(1), 116–125.

Akerlof, G. A., Kranton, R. E. (2000). Economics and identity. *Quarterly journal of Economics*, 715-753.

Akerlof, G. A., Kranton, R. E. (2002). Identity and schooling: Some lessons for the economics of education. *Journal of Economic Literature* 40: 1167-1201.

Akerlof, G. A., & Kranton, R. E. (2005). Identity and the Economics of Organizations. *Journal of Economic Perspectives* 19(1): 9-32.

Alesina, A., La Ferrara, E. (2005). Ethnic Diversity and Economic Performance. *Journal of Economic Literature* 43: 721-761.

Alesina, A., Baqir, R., Easterly, W. (1999). Public Goods and Ethnic Divisions, *Quarterly Journal of Economics* 114: 1243-1284.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth* 8: 155-194.

Alesina, A., Zhuravskaya, E. (2011). Segregation and the Quality of Government in a Cross Section of Countries. *American Economic Review* 101(5): 1872-1911.

Alexander, M., Christia, F. (2011). Context modularity of human altruism. *Science* 334(6061): 1392-1394.

Allport, G. W. (1954). The nature of prejudice. Cambridge, MA; Addison-Wesley.

Banerjee, A., Iyer, L., Somanathan, R. (2005). History, Social Divisions and Public Goods in Rural India. *Journal of the European Economic Association* 3(2-3): 639-647.

Bénabou, R., Tirole, J. (2011). Identity, Morals and Taboos. *Quarterly Journal of Economics* 126(2): 805-855.

Benjamin, D. J., Choi, J. J., & Fisher, G. (2016). Religious identity and economic behavior. Review of Economics and Statistics, 98(4), 617-637.

Bernhard, H., Fehr, E., Fischbacher, U. (2006a). Group affiliation and altruistic norm enforcement. *American Economic Review* 96(2): 217–221.

Bernhard, H., Fehr, E., Fischbacher, U. (2006b). Parochial altruism in humans. Nature 442: 912.

Bhattacharya, P. (2016). Inter-team contests with power differential. *Journal of Economic Behavior & Organization*, 132, 157-175.

Blanco, M., Engelmann, D., Koch, A.K., Normann, H.T. (2010). Belief elicitation in experiments: is there a hedging problem? *Experimental Economics* 13: 412-438

Brandts, J., Cooper, D. J., & Fatás, E. (2007). Leadership and overcoming coordination failure with asymmetric costs. *Experimental Economics*, 10(3), 269-284.

Brandts, J., Cooper, D. J., Fatás, E., & Qi, S. (2016). Stand by me: Help, heterogeneity and commitment in experimental coordination games. *Management Science*, 62, 2916-36.

Brañas-Garza, P., & Espinosa, M.P. (2011). Unravelling Public Good Games. *Games* 2(4), 434-451.

Brookins, P., Lightle, J. P., & Ryvkin, D. (2015). An experimental study of sorting in group contests. *Labour Economics*, *35*, 16–25.

Buchan, N. R., Brewer, M. B., Grimalda, G., Wilson, R. K., Fatás, E., & Foddy, M. (2011). Global social identity and global cooperation. *Psychological science*, *22*(6), 821-828.

Buchan, N. R., Fatás, E., & Grimalda, G. (2012). Connectivity and cooperation. *The Oxford Handbook of Economic Conflict Resolution, New York, NY: Oxford University Press Inc*, 155-79.

Buchan, N.R., Grimalda, G., Wilson, R., Brewer, M., Fatás, E., Foddy, M. (2009). Globalization and human cooperation. *Proceeding of the National Academy of Sciences* 106(11): 4138–4142.

Charness, G., Rigotti, L., Rustichini, A. (2007). Individual behavior and group membership. *American Economic Review* 97: 1340-1352.

Chen, H., & Lim, N. (2017). How Does Team Composition Affect Effort in Contests? A Theoretical and Experimental Analysis. *Journal of Marketing Research*, 54(1), 44-60.

Chen, Y., Li, S.X., 2009. Group Identity and Social Preferences**.** *American Economic Review* 99(1): 431-457.

Chuah, S.-H., Hoffmann, R., Ramasamy, B., Tan, J. H.W. (2014). Religion, ethnicity and cooperation: An experimental study. *Journal of Economic Psychology* 45: 33-43.

CIS (Centro de Investigaciones Sociológicas). (2007). La Identidad Nacional en España, report 2667, www.cis.es

Corcuera, J. (1979). Orígenes, ideología y organización del nacionalismo vasco (1876-1904), Madrid, Siglo XXI.

Criado, H., Herreros, F, Miller, L., Ubeda, P. (2015). Ethnicity and trust: A multifactorial experiment. *Political Studies* 63 (1): 131-152.

Criado, H., Herreros, F, Miller, L., Ubeda, P. (2018). The unintended consequences of political mobilization on trust: The case of the secessionist process in Catalonia. *Journal of Conflict Resolution* 62 (2): 231-253.

Croson, R. (2000). Thinking like a game theorist: Factors affecting the frequency of equilibrium play. *Journal of Economic Behavior & Organization* 41: 299-314.

Desmet, K., Breton, M., Ortuño-Ortín, I., Weber, S. (2011). The stability and breakup of nations: a quantitative analysis. *Journal of Economic Growth* 16(3): 183-213.

Desmet, K., Ortuño-Ortín, I., Wacziarg, R. (2012). The political economy of linguistic cleavages. *Journal of Development Economics* 97(2): 322-338.

Desmet, K., Ortuño-Ortín, I., Weber, S. (2009). Linguistic Diversity and Redistribution. *Journal of the European Economic Association* 7(6): 1291-1318.

Easterly, W. (2001). Can institutions resolve ethnic conflict? *Economic Development and Cultural Change* 49(4): 687-706.

Easterly, W., Levine, R. (1997). Africa's Growth Tragedy: Policies and Ethnic Divisions. *Quarterly Journal of Economics* 112 (4): 1203-1250.

Eckel, C.C., Fatás, E., Wilson, R. (2010). Cooperation and Status in Organizations. *Journal of Public Economic Theory* 12(4): 737-762.

Eckel, C.C., Grossman, P.J. (2005). Managing Diversity by Creating Team Identity. *Journal of Economic Behavior and Organization* 58(3): 371-392.

Fatás, E., Heap, S. P. H., Arjona, D. R. (2018). Preference conformism: An experiment. *European Economic Review*, *105*, 71-82

Fearon, J. (2003). Ethnic and Cultural Diversity by Country. *Journal of Economic Growth* 8: 195-222.

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

Gachter, S., Renner, E. (2010). The effects of (incentivized) belief elicitation in public goods experiments. *Experimental Economics* 13: 364-377.

Gaibulloev, K., Sandler, T. (2008). Growth Consequences of Terrorism in Western Europe. *Kyklos* 61(3): 411-424.

Gaibulloev, K., Sandler, T. (2009). The Impact of Terrorism and Conflicts on Growth in Asia. *Economics and Politics* 21(3): 359-383.

Gardeazabal, J. (2011). Linguistic Polarization and Conflict in the Basque Country. *Public Choice* 149(3/4): 405-425.

Goette, L., Huffman, D., Meier, S. (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.

Habyarimana, J., Humphreys, M., Posner, D.N., Weinstein, J.M. (2007). Why Does Ethnic Diversity Undermine Public Goods Provision? *American Political Science Review* 101(4): 709-725.

Hargreaves Heap, S. P., Ramalingam, A., Ramalingam, S., & Stoddard, B. V. (2015). "Doggedness" or "disengagement"? An experiment on the effect of inequality in endowment on behaviour in team competitions. *Journal of Economic Behavior & Organization*, *120*, 80–93.

INE (Spanish National Statistics Institute (http,//www.ine.es/), report 2667.

Keizer, K., Lindenberg, S., Steg, L. (2008) The Spreading of disorder. Science 322: 1681-1685.

Keuschnigg, M., Schikora, J. (2014). The Dark Side of Leadership: An Experiment on Religious Heterogeneity and Cooperation in India. *Journal of Socio-Economics* 48: 19-26.

Khwaja, A.I. (2009). Can good projects succeed in bad communities? *Journal of Public Economics* 93(7-8): 899-916.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., Vishny, R. (1999). The Quality of Government. *Journal of Law, Economics and Organization* 15 (1): 222-279.

Laitin, D. (1977). Politics, Language, and Thought: The Somali Experience. Chicago: University of Chicago Press.

Lee, T, and Pérez, E.O. (2014). The Persistent Connection Between Language-of-Interview and Latino Political Opinion. *Political Behavior* 36(2): 401-425

Li, S.X. (2010). Social Identities, Ethnic Diversity, and Tax Morale. *Public Finance Review* 38(2): 146-177.

Miguel, E., Gugerty, M.K. (2005). Ethnic Diversity, Social Sanctions and Public Goods in Kenya. *Journal of Public Economics* 89: 2325-2368.

Ministerio del Interior (2011). Election results, http://www.infoelectoral.interior.es/min/busquedaAvanzadaAction.html;jsessionid=11078856BC6736A57DE95477E97AF5E3.app2

Montalvo, J.G., Reynal-Querol, M. (2005). Ethnic diversity and economic development. *Journal of Development Economics* 76 (2): 293-323.

Neary, H. M. (1997). Equilibrium Structure in an Economic Model of Conflict. *Economic Inquiry* 35(3): 480-494.

Pérez, E. O. (2015). The Language-Opinion Connection. In R. Michael Alvarez and Lonna Atkeson, eds., *The Oxford Handbook of Polling and Polling Methods.* New York: Oxford University Press

Pérez, E. O. (2016). Rolling off the Tongue into the Top-of-the-Head: Explaining Language Effects on Public Opinion. *Political Behavior* 38(3): 603-634.

Pérez, E.O., and Tavits, M. (2017). Language Shapes People's Time Perspective and Support for Future-Oriented Policies. *American Journal of Political Science* 61(3): 715-727.

Sääksvuori, L., Mappes, T., & Puurtinen, M. (2011). Costly punishment prevails in intergroup conflict. *Proceedings of the Royal Society of London B: Biological Sciences*, *278*(1723), 3428-3436.

Santalucía, V. M., Martínez, L. S., de Guevara, J. F., García, F. P. (2005). La medición del capital social: una aproximación económica. *Fundación BBVA*

Shih, M., Pittinsky, T. L., Ambady, N. (1999). Stereotype Susceptibility: Identity Salience and Shifts in Quantitative Performance. *Psychological Science* 10(1): 80-83.

Singelis, T.M. (1994). The measurement of independent and interdependent self-construals. *Personality and Social Psychology Bulletin* 20: 580-591.

Singelis, T.M., Triandis, H.C., Bhawuk, D.S., Gelfand, M. (1995). Horizontal and vertical dimensions of individualism and collectivism: A theoretical and measurement refinement. *Cross-Cultural Research* 29*,* 240-275.

Tajfel, H., Turner, J. C. (1979). An Integrative Theory of Intergroup Conflict. In W. G. Austin & S. Worchel (Eds.). *The Social Psychology of Intergroup Relations*. Monterey, CA: Brooks-Cole.

Tanaka, T., Camerer, C.F. (2009). *Status and Ethnicity in Vietnam: Evidence from Experimental Games.* In: Social Computing and Behavioral Modeling. Springer, New York: 217-218.

Wagner, J. A. (1995). Studies of individualism-collectivism: Effects on cooperation in groups. *Academy of Management Journal 38*(1): 152-173.

Weng, Q., & Carlsson, F. (2015). Cooperation in teams: The role of identity, punishment, and endowment distribution. *Journal of Public Economics*, *126*, 25-38.

Tables and figures

Figure 1: Social Returns and MPCR

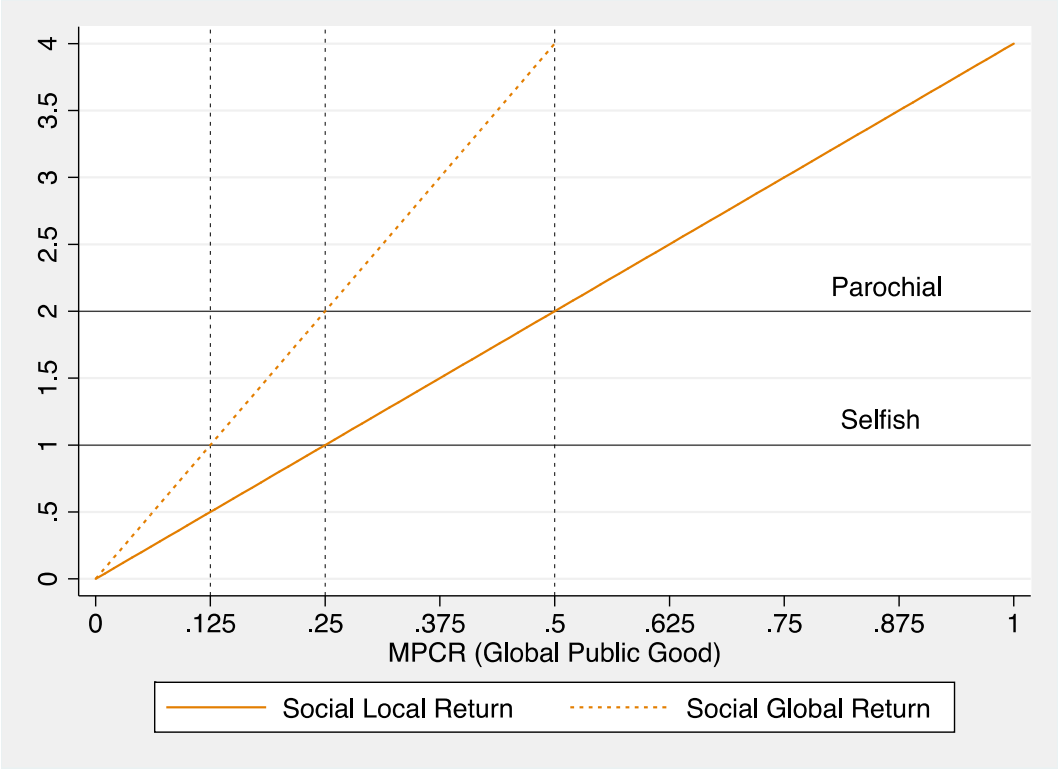


Table 1: Experiment I

Sessions	Language	Location	Subjects	Sections/Groups	Observations
Homogeneous	Spanish	Valencian Community	48	6/12	960
	Spanish	Basque Country	56	7/14	1120
	Catalan	Valencian Community	48	6/12	960
	Basque	Basque Country	56	7/14	1120
Mixed	Spanish/Catalan	Valencian Community	48	6/12	960
	Spanish/Basque	Basque Country	56	7/14	1120
Total	6	2	312	39/78	6240

Table 2: Average contribution to the local and global public good
(Experiment I, by location and language)

	Homogeneous		**Mixed**	
	Spanish	Catalan/Basque	All	
			Spanish	Cat/Bas
Global contribution				
Valencia	16.37 (22.51)	13.42 (20.85)	20.67	
			21.15 (28.01)	20.19 (28.91)
Bilbao	24.65 (24.49)	21.16 (27.29)	16.79	
			15.72 (20.66)	17.86 (22.61)
Local contribution				
Valencia	13.96 (19.48)	11.94 (21.93)	12.13	
			11.86 (20.44)	12.39 (20.85)
Bilbao	19.46 (21.19)	20.68 (28.23)	18.27	
			15.52 (21.31)	20.99 (24.04)

Standard deviations in parentheses. Results from Mann –Whitney non parametric test:
Valencia:
Local Homogeneous Spanish vs. Homogeneous Catalan: z=-0.320, p-value=0.7488
Local Homogeneous vs. Mixed: z=-0.234, p-value=0.8148
Global Homogeneous Spanish vs. Homogeneous Catalan: z=-0.961, p-value=0.3367
Global Homogeneous vs. Mixed: z=-1.218, p-value=0.2234
Bilbao:
Local Homogeneous Spanish vs. Homogeneous Basque: z=1.086, p-value=0.2774
Local Homogeneous vs. Mixed: z=0.075, p-value=0.9405
Global Homogeneous Spanish vs. Homogeneous Basque: z=-0.575, p-value=0.5663
Global Homogeneous vs. Mixed: z=1.903, p-value=0.0570

Figure 2a: Contribution to the global public good
(Experiment I, by location and treatment)

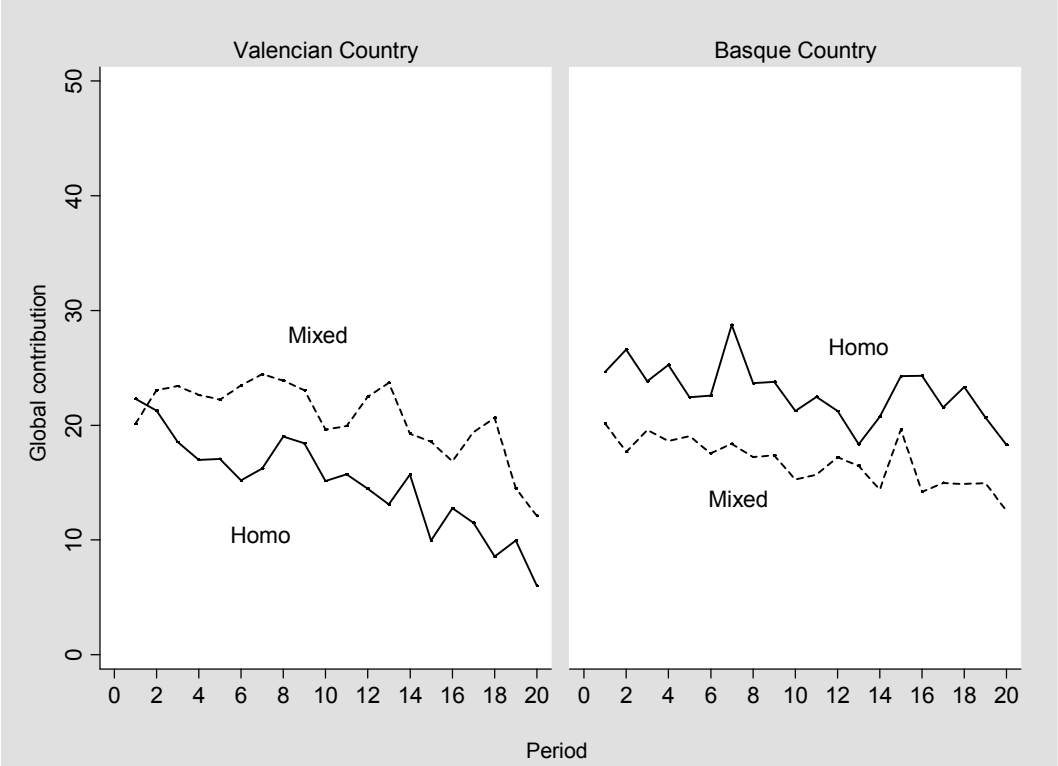


Figure 2b: Contribution to the local public good

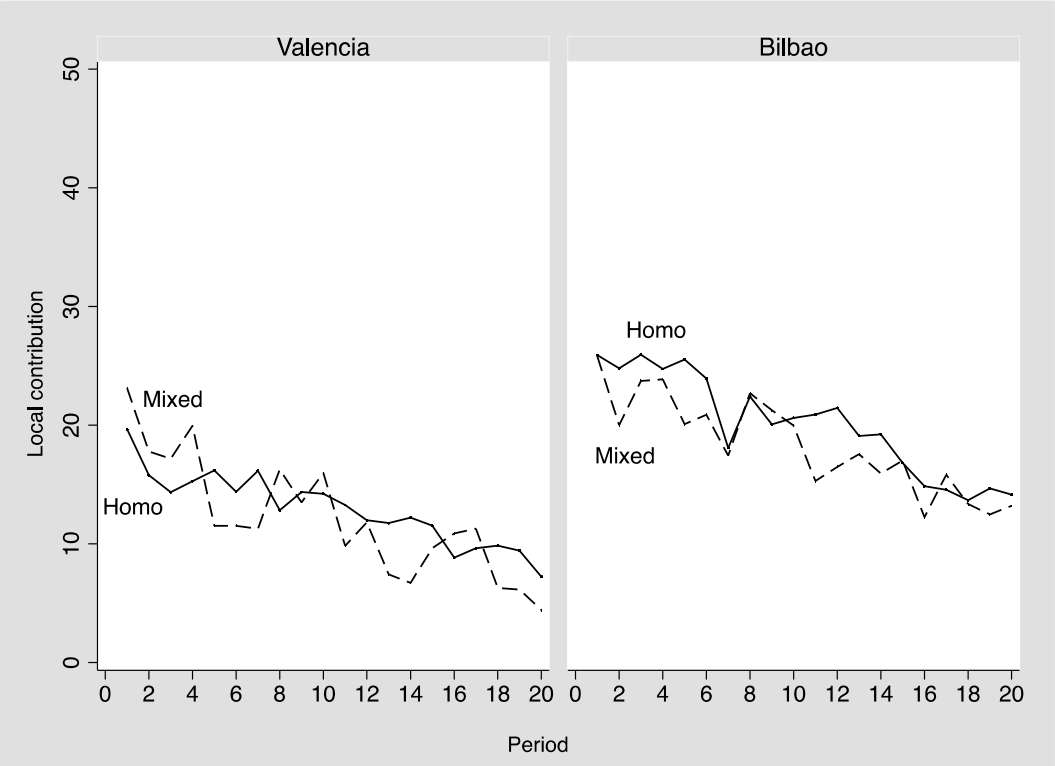


Figure 2c: Contribution to the private account

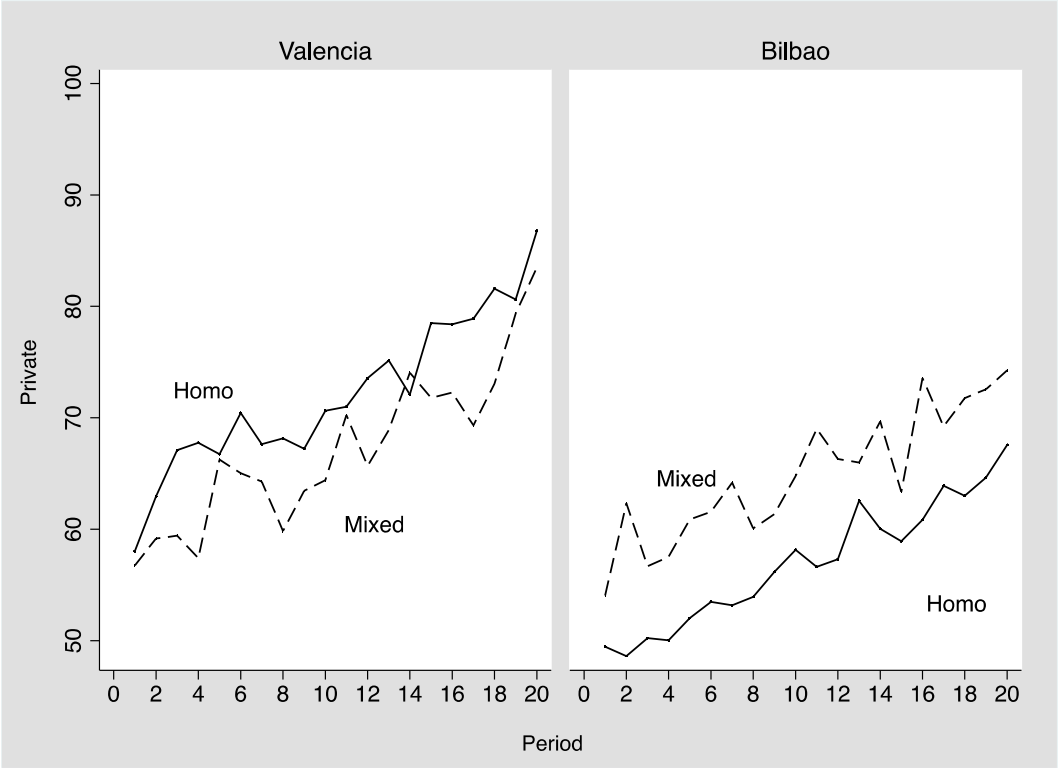


Table 3a: Contribution to the global public good
(Experiment I, panel data, treatment effects)

	Valencia	Valencia	Bilbao	Bilbao	Both
	(1)	(2)	(3)	(4)	(5)
Mixed	5.808	4.708	-6.796**	-10.353***	5.828
	(4.012)	(5.757)	(3.100)	(3.890)	(4.015)
BC	-	-	-	-	9.894***
					(2.477)
Language	-	-3.037	-	-2.867	-
		(2.928)		(3.902)	
*Mixed*BC*	-	-	-	-	-12.664**
					(5.037)
*Mixed*Language*	-	2.075	-	7.161	-
		(4.766)		(6.147)	
Period	-0.663***	-0.663***	-0.423***	-0.423***	-0.536***
	(0.096)	(0.096)	(0.093)	(0.093)	(0.069)
Local PG	-0.179***	-0.180***	-0.251***	-0.251***	-0.225***
	(0.051)	(0.051)	(0.042)	(0.042)	(0.033)
Constant	8.375	9.860	26.326*	29.889**	12.370
	(8.553)	(8.525)	(14.947)	(14.030)	(8.452)
Control variables	YES	YES	YES	YES	YES
Linear combination of estimators					
*Mixed+Mixed*Language*	-	6.782**	-	-3.192	-
		(3.228)		(4.724)	
*Mixed+Mixed*BC*	-	-	-	-	-6.837**
					(3.070)
*BC+Mixed*BC*	-	-	-	-	-2.771
					(4.386)
Observations	2880	2880	3360	3360	6240

Note: Global contributions are measured as the number of ECUS participants contribute to the global public good game. The explanatory variables used in the different models are: *Mixed*=1 if session is Mixed; =0 if session is Homogeneous; *BC=1 if contributions are made by participants in the Basque Country*; =0 if made by participants in the Valencian Community; *Language* =1 if the language spoken by participants was Catalan or Basque, =0 if Spanish; *Period* takes values from 1 to 20 indicating the round; the control variables are individual scores for *Individualism* and *Collectivism* in the Singelis scale (see the Appendix) and *Local PG*: subject's contribution to the local public good. Standard errors in parentheses. Panel data with random effects and cluster at the section level. *** p<0.01, ** p<0.05, * p<0.1

Table 3b: Contribution to the global public good
(Experiment I, panel data, treatment effects)

	Valencia	Valencia (Spanish)	Valencia (Catalan)	Bilbao	Bilbao (Spanish)	Bilbao (Basque)
Mixed	10.696	10.573	11.808	-6.536	-11.779*	-2.005
	(7.991)	(7.018)	(12.963)	(5.458)	(6.447)	(7.689)
Selfish	-0.014	-0.025	-0.006	0.016	-0.051	0.057
	(0.024)	(0.024)	(0.042)	(0.041)	(0.086)	(0.036)
Private	-0.559***	-0.576***	-0.539***	-0.569***	-0.569***	-0.568***
	(0.055)	(0.073)	(0.082)	(0.024)	(0.035)	(0.032)
Selfish*Mixed	-0.019	-0.050	-0.003	0.004	0.090	-0.074
	(0.035)	(0.035)	(0.082)	(0.047)	(0.093)	(0.055)
Private*Mixed	-0.101	-0.064	-0.134	0.069	0.082	0.060
	(0.078)	(0.095)	(0.108)	(0.058)	(0.061)	(0.097)
Period	0.098	0.038	0.152	0.238***	0.267**	0.209**
	(0.081)	(0.115)	(0.096)	(0.069)	(0.105)	(0.099)
Constant	44.538***	48.020***	42.052***	43.904***	51.711***	40.755***
	(7.995)	(9.517)	(12.931)	(7.944)	(8.121)	(10.658)
Control variables	YES	YES	YES	YES	YES	YES
Linear combination of estimators						
	*Mixed+Selfish*Mixed+ Private*Mixed*					
	10.576	10.459	11.671	-6.463	-11.607*	-2.019
	(7.894)	(6.938)	(12.809)	(5.387)	(6.359)	(7.578)
Observations	2880	1440	1440	3360	1680	1680

Note: Global contributions are measured as the number of ECUS participants contribute to the global public good game. The explanatory variables used in the different models are: *Mixed*=1 if session is Mixed; =0 if session is Homogeneous; *Selfish*: amount subjects take in their private account in the first period; *Private*: amount subjects keept in the private account in each period; the interaction terms Selfish*Mixed and Selfish*Private; *Period* takes values from 1 to 20 indicating the round; the control variables are individual scores for *Individualism* and *Collectivism* in the Singelis scale (see the Appendix). Standard errors in parentheses. Panel data with random effects and cluster at the section level. *** p<0.01, ** p<0.05, * p<0.1

Table 3c: Contribution to the local public good
(Experiment I, panel data, treatment effects)

	Valencia	Valencia (Spanish)	Valencia (Catalan)	Bilbao	Bilbao (Spanish)	Bilbao (Basque)
Mixed	-10.696	-10.573	-11.808	6.536	11.779*	2.005
	(7.991)	(7.018)	(12.963)	(5.458)	(6.447)	(7.689)
Selfish	0.014	0.025	0.006	-0.016	0.051	-0.057
	(0.024)	(0.024)	(0.042)	(0.042)	(0.086)	(0.036)
Private	-0.441***	-0.424***	-0.461***	-0.431***	-0.431***	-0.432***
	(0.056)	(0.073)	(0.083)	(0.024)	(0.035)	(0.032)
Selfish*Mixed	0.019	0.050	0.003	-0.004	-0.090	0.074
	(0.035)	(0.035)	(0.082)	(0.047)	(0.093)	(0.055)
Private*Mixed	0.101	0.064	0.134	-0.069	-0.082	-0.060
	(0.078)	(0.095)	(0.108)	(0.058)	(0.061)	(0.099)
Period	-0.098	-0.038	-0.152	-0.239***	-0.267**	-0.210**
	(0.082)	(0.115)	(0.096)	(0.069)	(0.105)	(0.099)
Constant	55.462***	51.980***	57.948***	56.096***	48.289***	59.245
	(7.995)	(9.518)	(12.931)	(7.944)	(8.132)	(10.658)
Control variables	YES	YES	YES	YES	YES	YES
Linear combination of estimators						
			*Mixed+Selfish*Mixed+ Private*Mixed*			
	-10.576	-10.459	-11.671	6.463	11.607*	2.019
	(7.894)	(6.938)	(12.809)	(5.387)	(6.359)	(7.578)
Observations	2880	1440	1440	3360	1680	1680

Note: Local contributions are measured as the number of ECUS participants contribute to the local public good game. The explanatory variables used in the different models are: *Mixed*=1 if session is Mixed; =0 if session is Homogeneous; *Selfish*: amount subjects take in their private account in the first period; *Private*: amount subjects keept in the private account in each period; the interaction terms Selfish*Mixed and Selfish*Private; *Period* takes values from 1 to 20 indicating the round; the control variables are individual scores for *Individualism* and *Collectivism* in the Singelis scale (see the Appendix). Standard errors in parentheses. Panel data with random effects and cluster at the section level. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Earnings
(Experiment I, panel data analysis, treatment variables)

	Valencia	Valencia	Bilbao	Bilbao	Both
	(1)	(2)	(3)	(4)	(5)
Mixed	17.463	11.832	-20.399**	-26.795**	17.223
	(12.565)	(15.396)	(9.900)	(11.780)	(12.461)
BC	-	-	-	-	31.901***
					(7.891)
Language	-	-9.415	-	-9.017	-
		(9.557)		(12.541)	
*Mixed*BC*	-	-	-	-	-38.097**
					(15.877)
*Mixed*Language*	-	10.897	-	13.010	-
		(10.834)		(12.351)	
Period	-2.243***	-2.243***	-1.429***	-1.429***	-1.805***
	(0 .257)	(0.258)	(0.259)	(0.259)	(0.193)
Constant	144.647***	149.695	197.281	206.216***	153.958***
	(21.291)	(22.237)	(35.313)	(34.680)	(20.418)
Control variables	YES	YES	YES	YES	YES
Linear combinations:					
*Mixed+Mixed*Language*	-	22.730*	-	-13.785	-
		(11.725)		(11.447)	
*Mixed+Mixed*BC*	-	-	-	-	-20.874**
					(9.772)
*BC+Mixed*BC*	-	-	-	-	-6.196
					(13.698)
Observations	2880	2880	3360	3360	6240

Note: Performance is measured as the number of ECUS participants earn in the experiment. The explanatory variables used in the different models are: *Mixed*=1 if session is *Mixed*; =0 if session is Homogeneous; *BC=1 if contributions are made by participants in the Basque Country*; =0 if made by participants in the Valencian Community; *Language* =1 if the language spoken by participants was Catalan or Basque, =0 if Spanish; *Period* takes values from 1 to 20 indicating the round; the control variables are individual scores for *Individualism* and *Collectivism* in the Singelis scale (see the Appendix). Standard errors in parentheses. Panel data with random effects and cluster at the section level. *** p<0.01, ** p<0.05, * p<0.1

Table 5a: Conditional Cooperation - Groups

	(1) Bilbao - Homo	(2) Bilbao - Mixed	(3) Valencia - Homo	(4) Valencia - Mixed
$Contribution_{I,t-1}$	0.473***	0.505***	0.488***	0.523***
	(0.0891)	(0.0766)	(0.0626)	(0.115)
$Contribution_{J,t-1}$	0.0825*	0.0527	0.0975**	0.249***
	(0.0501)	(0.0484)	(0.0455)	(0.0863)
Language	-2.768	1.528	-4.148	-2.434
	(3.185)	(5.127)	(4.882)	(12.41)
Period	-0.571**	-0.514	-1.100***	-0.922**
	(0.257)	(0.485)	(0.287)	(0.470)
Constant	48.88***	32.66***	36.22***	28.97*
	(6.161)	(11.59)	(10.82)	(15.21)
Observations	532	266	456	228
Groups	28	14	24	12

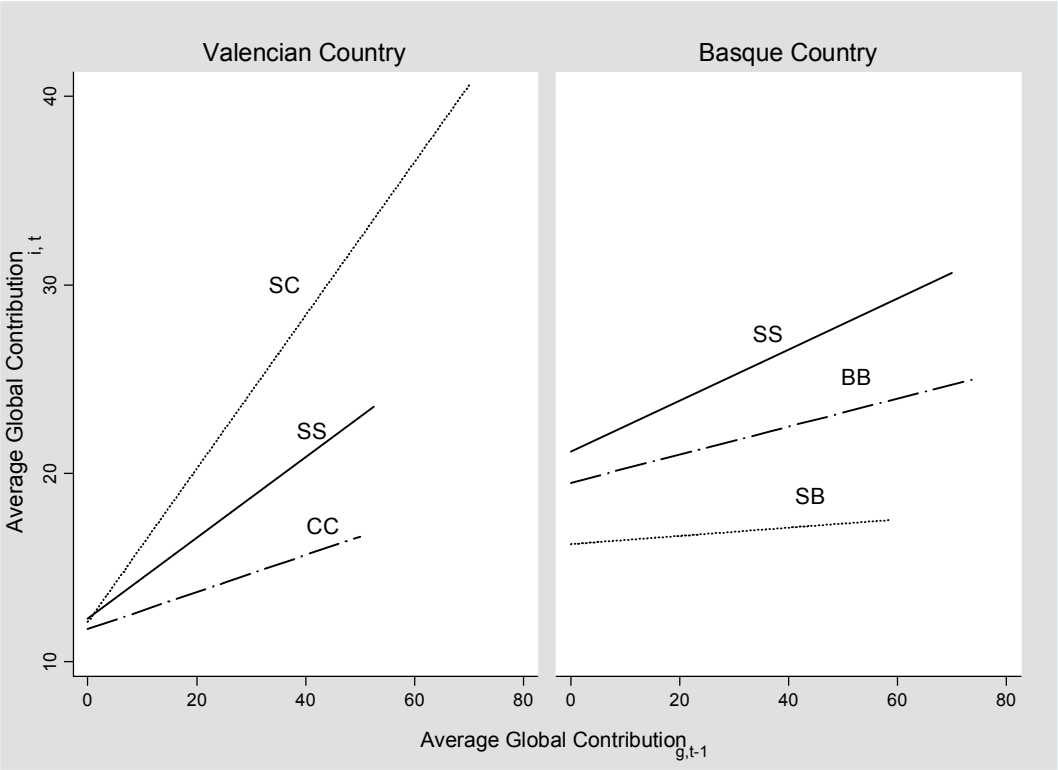
Note: Panel data with random effects at the group level, contributions are measured in ECUs, one per group and period. Standard errors clustered at the cohort level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 5b: Conditional Cooperation - Individuals

	Bilbao Homo	Bilbao Mixed	Valencia Homo	Valencia Mixed
$Contribution_{i,t-1}$	0.447***	0.469***	0.457***	0.521***
	(0.047)	(0.060)	(0.074)	(0.101)
$Contribution_{I-i,t-1}$	0.033	0.039	0.038	0.003
	(0.063)	(0.058)	(0.049)	(0.091)
$Contribution_{g,t-1}$	0.089*	0.049	0.105**	0.248***
	(0.052)	(0.055)	(0.044)	(0.078)
Period	-0.140**	-0.129	-0.267***	-0.230**
	(0.064)	0.121	(0.066)	(0.117)
Constant	11.201***	8.546***	8.226***	6.938***
	(2.072)	(3.069)	(2.054)	(2.775)
Observations	2128	1064	1824	912

Note: Panel data with random effects at the individual level, contributions are measured in ECUs. Standard errors clustered at the cohort level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Figure 3: Conditional cooperation
(Experiment I, global public good)



Note: SC stands for Mixed sessions with languages Spanish-Catalan; SB: Spanish-Basque; SS for Homogeneous sessions Spanish-Spanish; CC: Catalan-Catalan; BB: Basque-Basque.

Table 6: Decisions made by observers
(Experiment II)

1st order beliefs (about *participants* in Experiment I)

	Homogeneous		Mixed	
Decision:	**1**	**2**	**3**	**4**
Set 1 - Empirical	Spanish	Basque/Catalan	Spanish	Basque/Catalan
Set 2 - Normative	Spanish	Basque/Catalan	Spanish	Basque/Catalan

2nd order beliefs (about *observers* in Experiment II)

	Homogeneous		Mixed	
Decision:	**1**	**2**	**3**	**4**
Set 3 - Empirical	Spanish	Basque/Catalan	Spanish	Basque/Catalan
Set 4 - Normative	Spanish	Basque/Catalan	Spanish	Basque/Catalan

Table 7: 2nd order beliefs by observers' linguistic group
(Experiment II)

		Empirical expectations			**Normative expectations**			
		Own	IG Mixed	OG Mixed	Own	IG Mixed		OG Mixed
Valencia	Spanish	30.57	34.3	41.3**	44.67	41.37	××	48.77
		(17.14)	(20.07)	(25.28)	(27.27)	(29.93)		(29.93)
	Catalan	30.47	34.3**	38.57***	40.45	42.00	×××	48.00*
		(20.82)	(19.17)	(20.58)	(29.27)	(25.28)		(25.71)
	Obs	80	80	80	80	80		80
Bilbao	Spanish	37.37	36.97	38.47	40.29	43.42		45.16
		(18.88)	(21.72)	(23.01)	(27.85)	(30.78)		(31.68)
				++				+
	Basque	34.13	34.09	27.22	40.27	35.50		30.27**
		(21.85)	(20.55)	(22.21)	(28.44)	(31.41)		(30.52)
	Obs.	60	60	60	60	60		60

Standard deviations in parentheses

***, **, * significant differences with the Benchmark ($^{\varnothing}$<0.1014) using a Wilcoxon signed-rank test

×××,××,× significant differences between IG and OG Mixed expectations using a Wilcoxon signed-rank test

+++, ++, + significant differences between Spanish observers and Basque/Catalan observers using a Mann-Whitney rank sum test

All results at 1%, 5% and 10% significance level

Table 8: Empirical (descriptive) and normative (injunctive) norm adjustments (Experiment II)

		Spanish	Basque/Catalan	Spa-Mixed	Local-Mixed
Information disclosed	Valencia	16.37	13.42	21.15	20.19
(rounds 1-10)	Bilbao	24.65	21.16	15.72	17.86

	Observers	Spanish	Basque/Catalan	Spa-Mixed	Local-Mixed
Valencia – Empirical	Spanish	30.57	33.80	34.30	41.30
	Catalans	30.47	34.47	34.30	38.56
Adjustment	*Spanish*	*22.21*	*19.16*	*22.14*	*25.86*
		[0.0018]	*[0.0001]*	*[0.0007]*	*[0.0003]*
	Catalans	*21.25*	*23.08*	*25.84*	*25.93*
		[0.0084]	*[0.0001]*	*[0.0149]*	*[0.0007]*
Valencia - Normative	Spanish	44.68	45.05	41.38	48.78
	Catalans	40.45	46.53	42.00	48.00
Adjustment	*Spanish*	*35.90*	*31.70*	*34.05*	*37.45*
		[0.0084]	*[0.0045]*	*[0.0288]*	*[0.0054]*
	Catalans	*34.00*	*36.64*	*39.13*	*42.55*
		[0.0651]	*[0.0012]*	*[0.1722]*	*[0.0441]*
Bilbao – Empirical	Spanish	37.38	40.92	36.97	38.47
	Basque	34.14	33.23	34.09	37.23
Adjustment	*Spanish*	*28.30*	*26.01*	*19.55*	*22.20*
		[0.0065]	*[0.0008]*	*[0.0001]*	*[0.0008]*
	Basque	*27.48*	*31.75*	*22.34*	*16.63*
		[0.1483]	*[0.9095]*	*[0.0283]*	*[0.0389]*
Bilbao – Normative	Spanish	40.30	46.88	44.43	45.17
	Basque	40.27	33.91	35.50	30.27
Adjustment	*Spanish*	*41.90*	*34.61*	*30.55*	*33.59*
		[0.6265]	*[0.0109]*	*[0.0030]*	*[0.0190]*
	Basque	*32.17*	*29.97*	*28.18*	*24.39*
		[0.0816]	*[0.1706]*	*[0.1212]*	*[0.1597]*

p-values between brackets, corresponding to the outcome of Wilcoxon signed-rank tests comparing norms before and after getting information about the behavior of participants in rounds 1 to 10 of Experiment I

Figure 4: Out-group discrimination
(Experiment II, empirical and normative)

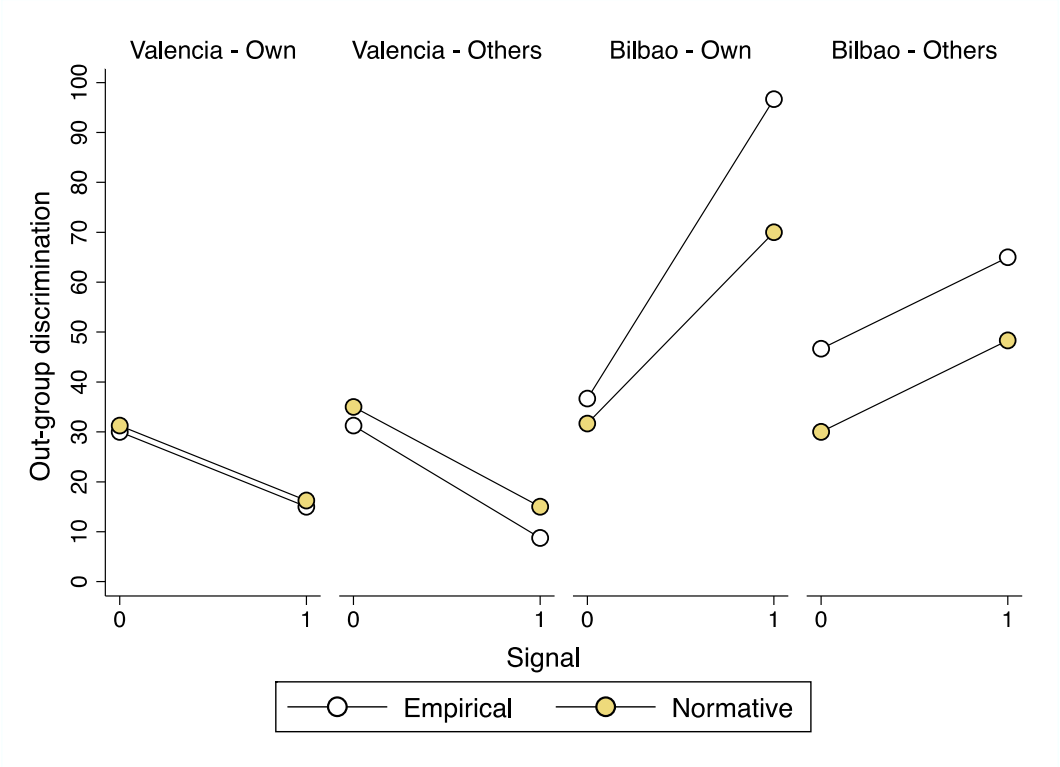


Table 9: Percentage of observers in Experiment II expecting out-group discrimination in the in-group (own) and the out-group (others).

Expectations	VLC	*p-value (2)*	BIO
Empirical Own B	30.00	*0.258*	36.67
p-value (1)	*0.0214*		*0.0000*
Empirical Own A	15.00	*0.000*	96.67
Normative Own B	31.25	*0.551*	31.67
p-value (1)	*0.0178*		*0.0001*
Normative Own A	16.25	*0.000*	70.00
Empirical Others B	31.25	*0.046*	46.67
p-value (1)	*0.0007*		*0.0354*
Empirical Others A	8.75	*0.000*	65.00
Normative Others B	35.00	*0.330*	30.00
p-value (1)	*0.0045*		*0.0261*
Normative Others A	15.00	*0.000*	48.33

Expectations	VLC	*p-value (2)*	BIO
Empirical Own B	30.00	*0.258*	36.67
p-value (1)	*0.5000*		*0.1316*
Empirical Others B	31.25	*0.046*	46.67
Normative Own B	31.25	*0.551*	31.67
p-value (1)	*0.3714*		*0.5000*
Normative Others B	35.00	*0.330*	30.00
Empirical Own A	15.00	*0.000*	96.67
p-value (1)	*0.1796*		*0.0000*
Empirical Others A	8.75	*0.000*	65.00
Normative Own A	16.25	*0.000*	70.00
p-value (1)	*0.5000*		*0.0073*
Normative Others A	15.00	*0.000*	48.33

Notes:

(1) Sign test for equality of matched pairs

(2) Fisher exact test

"B" after the name of a variable denotes before information disclosure and "A" indicates the same variable after information disclosure.

Appendix

Table A.1 Characteristics of the field

Socio-demographic variables	Basque Country	Valencian Community
Geographical segregation	No	No
School segregation	No	No
Annual expenditure per person (euros), 2011	13,774	10,140
GDP per capita 2011	31,288	20,583
Index of social capital 2004	700[1]	820[2]
Population (2012)	2,108,700	4,993,000
% Foreign population 2012	7.17	17.61
% Population (25 to 64 year old) with a college degree	45.0	29.5
Language (knowledge and use)		
Understands (Basque/Catalan)	55.5	94.7
Fluent (Basque/Catalan)	35.7	55.7
Uses mainly Spanish	64.7	60.5
Uses mainly Basque/Catalan	14.0	21.7
Which language was spoken at home when you were a child?		
Basque/Catalan	18.2	28.8
Spanish	68.7	60.8
Both	7.8	9.5
National identity		
Feel strongly identified with their country (Basque, Valencia) (9-10/10)	49.5	47.0
Feel strongly identified with Spain (9-10/10)	16.2	58.2
Prefer a state where the regions have the right to become independent	26.3	4.2
Although my official documents are Spanish, I do not consider myself Spanish (agree or strongly agree)	31.2	5.5
I feel only Spanish	5.3	12.3
I feel only Basque/Catalan	20.8	1.8
I feel more Spanish than Basque/Catalan	8.0	17.7
I feel more Basque/Catalan than Spanish	27.7	11.0
I feel equally Spanish than Basque/Catalan	32.2	55.0
Universities		
Names	Basque Country U	U of Valencia
Type	Public	Public
Number of campuses	3	4
Tuition fee per 10 hours, social sciences degree	€14.39	€14.96
Number of undergraduate students	36,377	38,576
Male students	16,462	14,148
Female students	19,081	22,637
% of Spanish students	92.31%	94.91%
Faculty members	3,849	5,316
QS ranking 2019 (World)	561-570	601-650
QS ranking 2019 (Spain)	15	19

Source: CIS (Center for Sociological Research) 2007. National Identity in Spain, report 2667, INE and Santalucia et al. (2005); [1]: fourth quartile; [2]: second-third quartile.

Source for Universities: Spanish Ministry of Education

Table A.2 Nationalism. Results from the regional elections 2011.

Basque Country	Valencian Community
EAJ-PNV (30.05%)	PP (48.53%)
Bildu-EA (25.45%)	PSOE (27.5%)
PSE-EE (16.34%)	**C.Compromis (7.03%)**
PP (13.53%)	**EUPV (5.79%)**
EB-B (3.19%)	UPyD (2.44%)
ARALAR (2.96%)	Verds (1.28)
Total Nationalist: 58.46%	**Total Nationalist: 12.82%**

Note: Nationalist parties in bold.
Source: Spanish Ministry for Internal Affairs (2011).

Figure A.1: Average local and global contributions by location and language

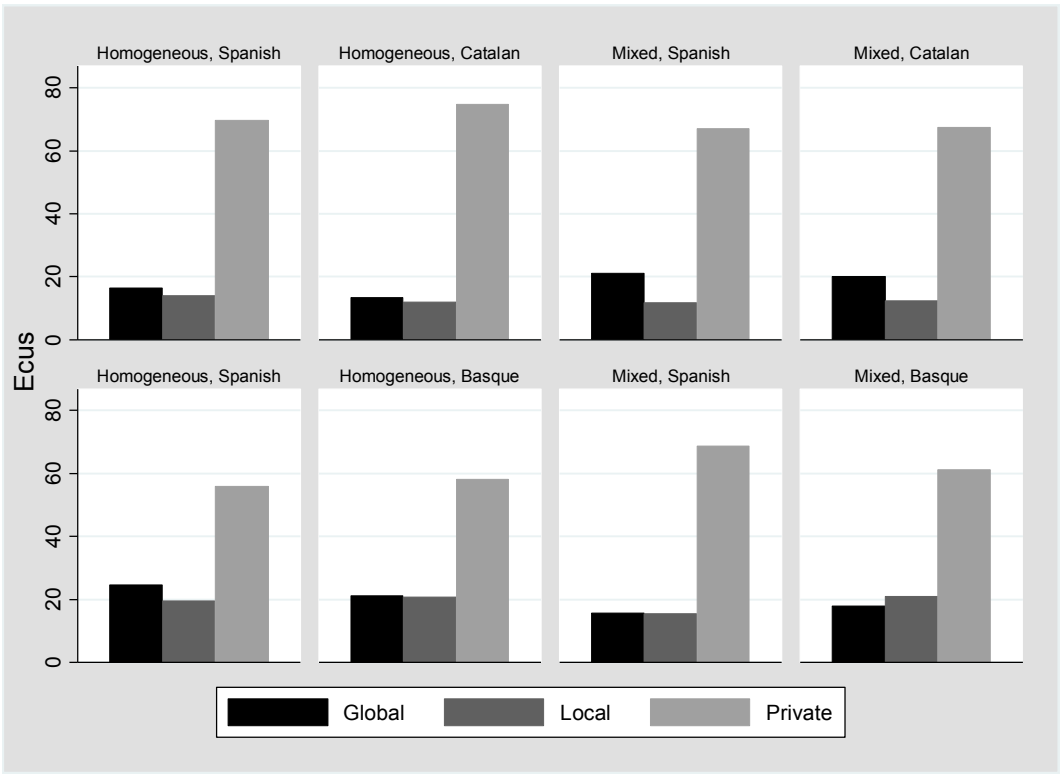
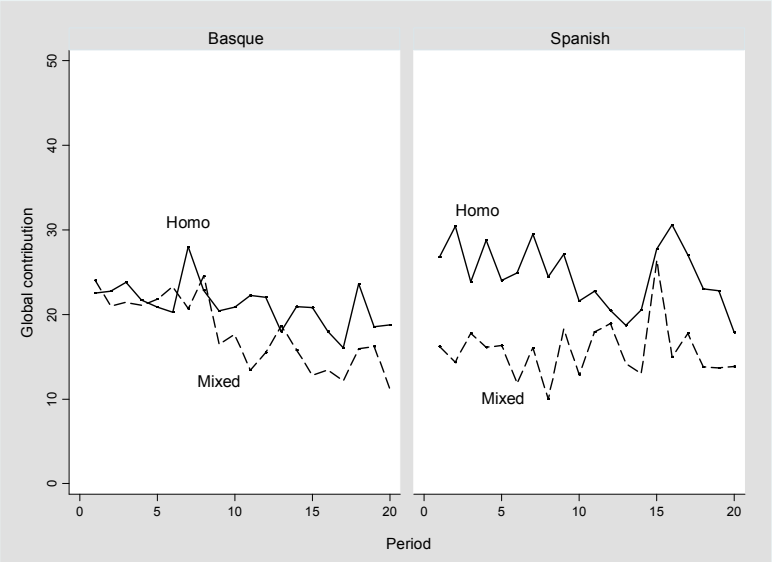


Figure A.2. Average global contribution by language and treatment

Basque Country



Valencian Community

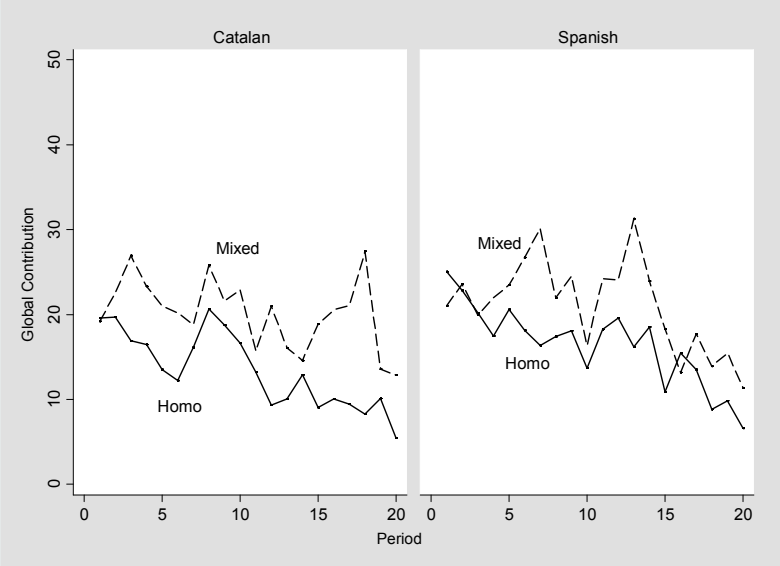
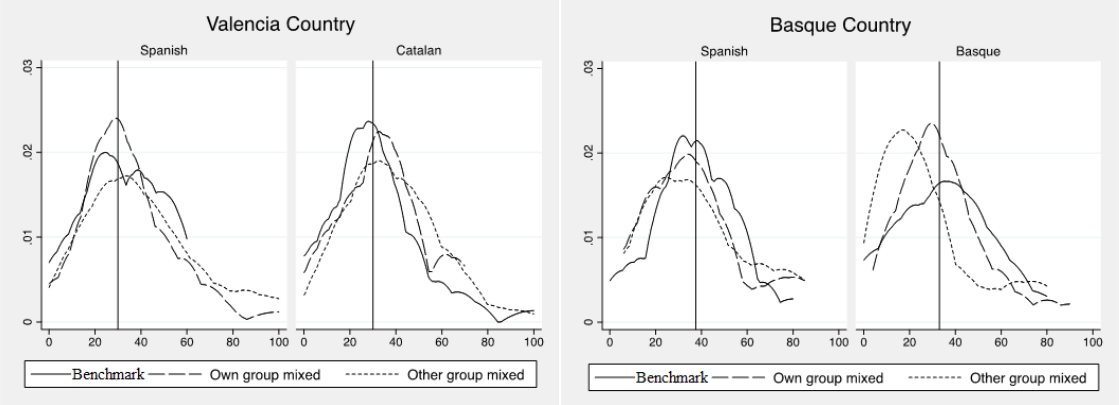


Figure A.3. Second order beliefs

Empirical



Normative

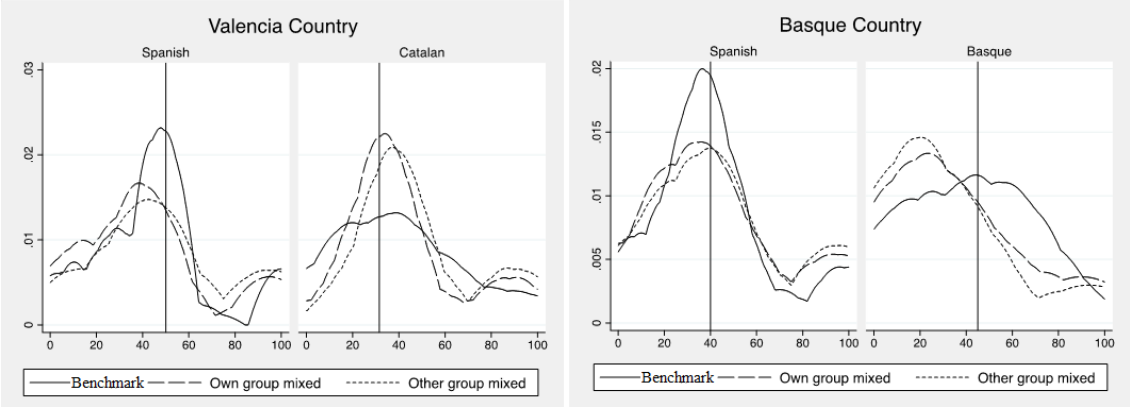
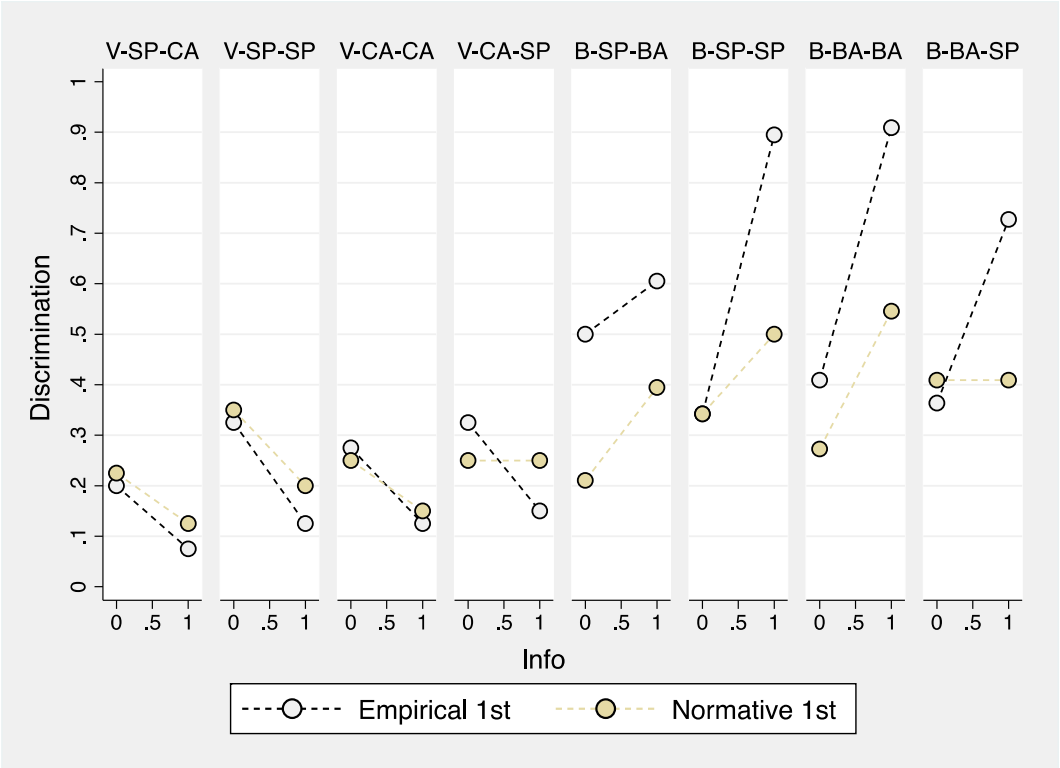
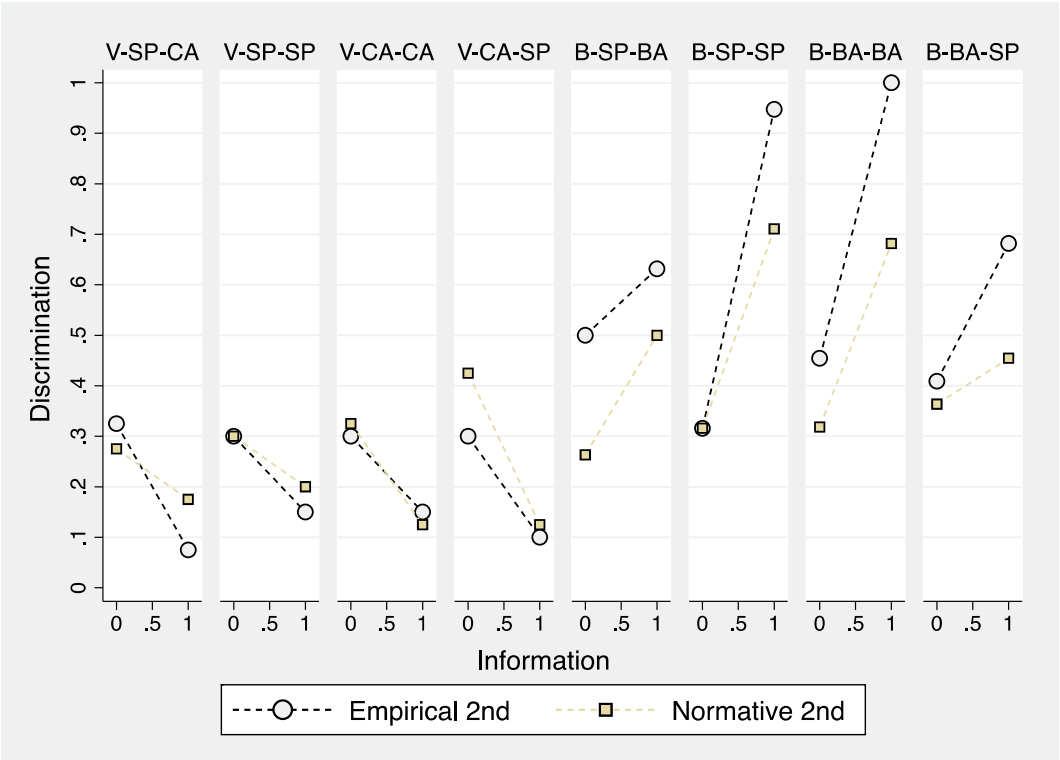


Figure A.4: Discriminatory expectations

(by location: V for Valencia and B for Bilbao;
by language of Observers: SP/CA/BA for Spanish/Catalan/Basque
by language of Participants: SP/CA/BA for Spanish/Catalan/Basque)





Measures of collectivism, individualism (Singelis, 1994)

Concerning the Singelis subjective-individualism-collectivism scale test, the distribution of scores for *individualism* was not the same for all linguistic groups (Spanish vs. Basque in the Basque Country z=2.274, p-value=0.0230; Spanish vs. Catalan in the Valencian Community z=1.402, p-value=0.1610). However, we do not find significant differences for *collectivism* (Spanish vs. Basque z=-1.584, p-value=0.1131; Spanish vs. Catalan z=-1.587, p-value=0.1124).

Individualism: Spaniards in the BC are not different from Spaniards in the VC (z=0.607, p-value=0.5440; t=0.3156, p-value=0.7528). We perform the test Spaniards vs. Catalans (z=1.484, p-value=0.1377; t=1.2997, p-value=0.1950) and Spaniards vs Basques (z=2.627, p-value=0.0086; t=2.2044, p-value=0.0285) by pooling the data from all the Spanish speaking subjects. Thus, Spaniards scored higher in individualism than Basques but Spaniards and Catalans are similar in this construct.

Collectivism: Spaniards in the BC are not different from Spaniards in the VC (z=-0.302, p-value=0.7623; t=-0.2639, p-value=0.7922). We perform the test Spaniards vs. Catalans (z=-1.706, p-value=0.0881; t=-1.5710 p-value(one-tail)=0.1176, p-value(two-tails)=0.0588) and Spaniards vs. Basques (z=-1.941, p-value=0.0523; t=-1.8761, p-value=0.0619) by pooling the data from all the Spanish speaking subjects . Thus, Spaniards scored lower in collectivism than Basques and Catalans.

Table A3. The effect of treatment variables in Global contributions

	Tobit		**Fractional multinomial model**	
Mixed	6.321	8.151**	0.054	0.066
	(5.602)	(3.993)	(0.052)	(0.041)
Location	14.824***	13.748***	0.082**	0.084***
	(4.393)	(3.142)	(0.033)	(0.022)
Language	-5.575	-	-0.032	-
	(4.617)		(0.032)	
Mixed*Location	-19.732**	-15.385***	-0.118***	-0.108***
	(7.656)	(5.435)	(0.043)	(0.034)
Mixed*Language	3.564	-	0.022	-
	(7.903)		(0.045)	
Location*Language	-2.323	-	0.004	-
	(6.217)		(0.044)	
Mixed*Location*Language	8.963	-	0.033	-
	(10.794)		(0.078)	
Period	-0.674***	-0.674***	-0.004***	-0.004***
	(0.064)	(0.064)	(0.001)	(0.001)
Constant	-8.207	-13.078	-1.770***	-1.959***
	(14.365)	(14.231)	(0.591)	(0.594)
Control variables	YES	YES	YES	YES
Observations	6240	6240	6240	6240

Standard errors in parentheses. Cluster at the section level. Tobit analysis is left-censored. *** p<0.01, ** p<0.05, * p<0.1

Explanatory variables. Mixed, dummy variable (1 in Mixed sessions); *Location*, dummy variable (1 in the BC); *Language*, dummy variable (1 if Catalan or Basque); *Period*, round 1 to 20; the control variables are individual scores for *Individualism* and *Collectivism* in the Singelis scale.

Table A4: Panel data analysis on the effect of treatment variables on Local contributions

	VC (1)	VC (2)	BC (3)	BC (4)	VC+BC (5)
Mixed	0.269	-0.993	-3.139	-6.542**	0.571
	(2.300)	(4.297)	(2.730)	(2.747)	(2.421)
BC	-	-	-	-	8.787***
					(2.540)
Language	-	-1.884	-	0.683	-
		(3.904)		(3.678)	
*Mixed*BC*	-	-	-	-	-3.529
					(3.609)
*Mixed*Language*	-	2.457	-	6.732	-
		(6.409)		(5.083)	
Period	-0.625***	-0.625***	-0.709***	-0.709***	-0.678***
	(0.069)	(8.006)	(0.066)	(0.066)	(0.048)
Global PG	-0.128***	-0.128***	-0.222***	-0.222***	-0.181***
	(0.031)	(0.031)	(0.037)	(0.037)	(0.026)
Constant	22.963***	24.004***	38.936*	40.036**	26.267**
	(8.836)	(8.006)	(20.095)	(20.311)	(10.649)
Control variables	YES	YES	YES	YES	YES
Linear combination of estimators					
*Mixed+Mixed*Language*	-	1.464	-	0.191	-
		(3.520)		(4.442)	
*Mixed+Mixed*BC*	-	-	-	-	-2.958
					(2.682)
*BC+Mixed*BC*	-	-	-	-	5.258**
					(2.627)
Observations	2880	2880	3360	3360	6240

Note: Local contributions are measured as the number of ECUS participants contribute to the local public good game. The explanatory variables used in the different models are: *Mixed*=1 if session is Mixed; =0 if session is Homogeneous; *BC=1 if contributions are made by participants in the Basque Country*; =0 if made by participants in the Valencian Community; *Language* =1 if the language spoken by participants was Catalan or Basque, =0 if Spanish; *Period* takes values from 1 to 20 indicating the round; the control variables are individual scores for *Individualism* and *Collectivism* in the Singelis scale (see the Appendix) and *Global PG* that is the number of ECUS participants contribute to the global public good game. Standard errors in parentheses. Panel data with random effect and cluster at the section level. *** p<0.01, ** p<0.05, * p<0.1

Supporting Online Material for

Linguistic Diversity and Out-Group Discrimination in Bilingual Societies

This PDF file includes:

I. Instructions for Part I of the experiment
II. Pre-experimental questionnaire.
III. Post-experimental questionnaire: Singelis test
IV. Instructions for Part II of the experiment

I. EXPERIMENTAL INSTRUCTIONS FOR PART I

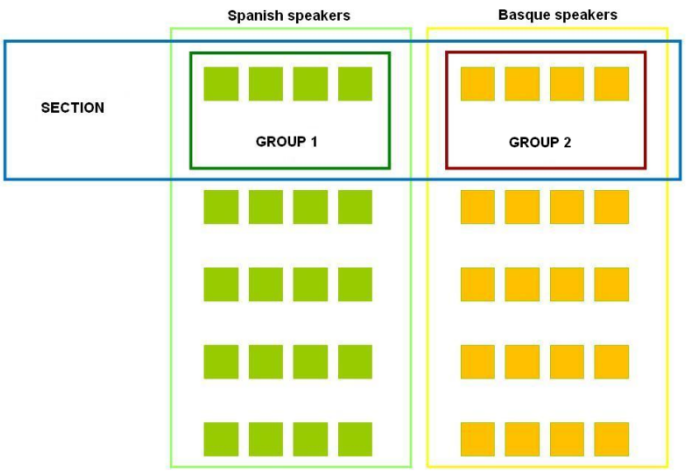
The aim of the experiment is to study how individuals make decisions in certain environments. Instructions are easy to follow and you can make some money if you follow them carefully. Money will be privately paid at the end of the experiment. Should you have any questions please raise your hand and an assistant will come to you. Any communication between you and other participants is strictly forbidden. If you do not follow this rule, you will be excluded from the experiment.

1. You will receive 2000 ECUs (experimental currency units) merely for taking part in this experiment that will be exchanged for Euros at the end of the experiment.

2. This experiment will be conducted in the two official languages of the UPV/EHU. Given that the participants choose the session language (Spanish or Basque), the participants in this session are divided in two laboratories. Only one language will be spoken in each laboratory. In today's experiment, the participants in this laboratory have chosen to do the experiment in Spanish while the participants in the other laboratory have chosen to do it in Basque.

3. The experiment consists of 20 rounds. You are a member of a section of 8 participants In each round. Each section of 8 participants is composed of two groups of 4 participants. You belong to one group of 4 participants. One group of 4 participants is composed of participants from this laboratory and the other group is composed of participants from the other laboratory. Both groups are connected by the net.

4. The composition of each group and each section is randomly determined at the beginning of the experiment and does not change during the experiment. You will never know the identities of the other group and section members. The following figure represents the distribution of participants in groups and sections in the two laboratories. [35]



5. Every participant gets an initial endowment of 100 ECUs (Experimental Currency Units) in each round. You must only decide how much of this amount you want to assign to the Group Account and how much you want to assign to the Section Account. The remainder will automatically be assigned to an Individual Account.

[35] For illustrative purposes, we represent members of a group sitting next to each other; but of course members of a group and a section were randomly matched. Subjects were aware of this design.

6. Your payoff from the Individual Account (IA) equals your assignment to the Individual Account and does not depend on the decisions of others.
7. The payoff from the Group Account (GA) and the Section Account (SA) depends on the amount of ECUs you have assigned to those accounts as well as the amount assigned by the members of your group and your section to the Group Account and Section Account in the following way:
    a. The sum of all assignments of your Group to the Group Account (GA) will multiplied by two and equally divided among the 4 members of the Group, regardless of your assignment to that Account.
    b. The sum of all assignments of your Section to the Section Account (SA) will be multiplied by four and equally divided among the 8 members of the Section, regardless of your assignment to that Account.

8. In summary, your earnings in one round will be determined as follows:

$$\text{Individual earnings} = \underset{\textit{100 ECU – my allocation to the GA and SA}}{\text{Payoff IA}} + \text{Payoff GA}\ \frac{\textit{2 x GA of my group}}{\textit{4}} + \text{Payoff SA}\ \frac{\textit{4 x SA of my section}}{\textit{8}}$$

9. A simple way to understand how your earnings in each round have been calculated is the following:
   Your earnings in a round are the sum of:
    a. The ECUs you decided not to allocate neither to the Group account nor to the Section Account (the amount of ECUs that goes to the Individual Account (IA)).
    b. 50% of the total amount assigned by the 4 members of your group to the Group Account **(2/4).**
    c. 50% of the total amount assigned by the 8 members of your section to the Section Account **(4/8).**

10. An example will allow you to better understand how your decisions will be converted into earnings at the end of the experiment. For simplicity purpose, let us focus on one of the groups of one section of 8 participants. We will call this group, Group 1 (the other 4 members of the Section belong to Group 2, located in the other laboratory).

Suppose that the members of Group 1 make the following **allocations:**

		GA of Group 1	*SA*	*IA*
i.	Participant A	100	0	0
ii.	Participant B	0	100	0
iii.	Participant C	0	0	100
iv.	Participant D	40	20	40

The members of Group 1 have allocated a total of 140ECUs to the GA (100+0+0+40) and 120 ECUs to the SA (0+100+0+20). Suppose that the participants from Group 2 have allocated a total of 40 ECUs to the SA. Then earnings for members of Group 1 are calculated as follows:

		Earning GA	*Earning SA*	*Earning IA*	*Total earnings*
i.	Participant A	70	80	0	**150**
ii.	Participant B	70	80	0	**150**
iii.	Participant C	70	80	100	**250**
iv.	Participant D	70	80	40	**190**

11. Note that the earning that each participant gets from the Group Account is the same (regardless of how much you allocate to that Account) and members of Group 2 also get 80 ECUs from the Section Account although they have allocated in total less than Group 1 to that Account. Thus, all the members of each group (and each section) will get always half of the total allocation made to the GA and SA, regardless of how much they have assigned to the Accounts. Remember that you are not able to assign to these two Accounts more than 100 ECUs by round.
12. After each round you will be given information about individual allocations of the members of your group to the GA and to the SA, as well as the allocations of the members of the other group to the SA. You will also get information about your earnings in each round (earnings from GA, SA and IA). The information of the allocation and earnings from the past rounds will appear on your computer screen. The following table shows the information that Participant A in the above example would observe:

	My allocations			My Group		The other Group	Section	Earnings		
Round	IA	GA	SA	GA	SA	SA	SA	IA	GA	SA	Total
1	0	100	0	140	120	40	160	0	70	80	150

13. At the end of the experiment, the sum of your individual earnings over the 20 rounds will be privately paid to you at the exchange rate of 400 ECUs=€1.

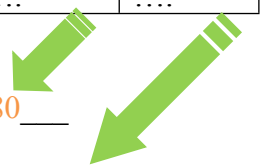
14. II. PRE-EXPERIMENTAL QUESTIONNAIRE

Questionnaire

Choose the correct answer and fill out the gaps. When you finish, please raise your hand and an assistant will check the answers. The assistant will inform you whether your answers are correct or not. If you have made any mistake the assistant will give back the questionnaire and you will have to fill it again. You cannot ask any question to the assistant about the correct answers. Any communication between you and other participants is forbidden. If you do not follow this rule, you will be excluded from the experiment.

The following example shows the decisions from participants of one group (Group 1) in a section and in one period. It also includes basic information about the decisions of the participants of the other group (Group 2) in the same section. Please fill in the gaps.

	Private Account (IA)	GA Group 1	SA Section
Participant 1	50	50	0
Participant 2	0	0	100
Participant 3	70	30	0
Participant 4	100	0	0
Total allocations of my group:		….	….



Total allocation of my group to the GA of Group 1: ___80___

Total allocation of my group to the GA of the Section: ___100___

Suppose that Group 2 allocates 80 ECUs to the SA.
Thus, the sum of the allocations of both groups to the SA is: __180____

The earnings that each participant in Group 1 gets from the GA of Group 1 is half of the value of all the allocations to that account: ___40___

The earnings that each participant in Group 1 gets from the SA is half of the value of all the allocations to that account:__90____

The following table shows the earnings of the members of Group 1. Please fill in the gaps.

	IA	GA G1	SA	Total Earnings
Participant 1	50	40	90	180
Participant 2	0			130
Participant 3	70			200
Participant 4	100			230

III. POST-EXPERIMENTAL QUESTIONNAIRE: SINGELIS TEST

INSTRUCTIONS

This is a questionnaire that measures a variety of feelings and behaviors in various situations. Listed below are a number of statements. Read each one as if it referred to you. Beside each statement write the number that best matches your agreement or disagreement. Please respond to every statement. Thank you.

1=STRONGLY DISAGREE	4=DON'T AGREE OR DISAGREE	5=AGREE SOMEWHAT
2=DISAGREE		6=AGREE
3=SOMEWHAT DISAGREE		7=STRONGLY AGREE

____1. I enjoy being unique and different from others in many respects.

____2. I can talk openly with a person who I meet for the first time, even when this person is much older than I am.

____3. Even when I strongly disagree with group members, I avoid an argument.

____4. I have respect for the authority figures with whom I interact.

____5. I do my own thing, regardless of what others think.

____6. I respect people who are modest about themselves.

____7. I feel it is important for me to act as an independent person.

____8. I will sacrifice my self-interest for the benefit of the group I am in.

____9. I'd rather say "No" directly, than risk being misunderstood.

____10. Having a lively imagination is important to me.

____11. I should take my parents' advice into consideration when making education/career plans.

____12. I feel my fate is intertwined with the fate of those around me.

____13. I prefer to be direct and forthright when dealing with people I've just met.

____14. I feel good when I cooperate with others.

____15. I am comfortable with being singled out for praise or rewards.

____16. If my brother or sister fails, I feel responsible.

____17. I often have the feeling that my relationships with others are more important than my own accomplishments.

____18. Speaking up during a class (or a meeting) is not a problem for me.

____19. I would offer my seat in a bus to my professor (or my boss).

____20. I act the same way no matter who I am with.

____21. My happiness depends on the happiness of those around me.

____22. I value being in good health above everything.

____23. I will stay in a group if they need me, even when I am not happy with the group.

____24. I try to do what is best for me, regardless of how that might affect others.

____25. Being able to take care of myself is a primary concern for me.

____26. It is important to me to respect decisions made by the group.

____27. My personal identity, independent of others, is very important to me.

____28. It is important for me to maintain harmony within my group.

____29. I act the same way at home as I do at school (or work).

____30. I usually go along with what others want to do, even when I would rather do something different.

To score the scale, add each subject's scores (1 to 7) for the independent items and divide by 15 to give the mean score of the items. Then, do the same for the interdependent items. ***Each subject receives two scores***: one for the strength of the independent self and one for the interdependent self.

Independent items: 1, 2, 5, 7, 9, 10, 13, 15, 18, 20, 22, 24, 25, 27, and 29.

Interdependent items: 3, 4, 6, 8, 11, 12, 14, 16, 17, 19, 21, 23, 26, 28, and 30.

INSTRUCTIONS FOR PART II IF THE EXPERIMENT

The aim of the experiment is to study how individuals make decisions in some environments. Instructions are easy to follow and you can make some money if you follow them carefully. Money will be privately paid at the end of the experiment. Should you have any questions please raise your hand and an assistant will come to you. Any communication between you and other participants is strictly forbidden. If you do not follow this rule, you will be excluded from the experiment.

1. This experiment will be conducted in the two official languages of the UPV/EHU. In today's experiment, the participants in this laboratory have chosen to do the experiment in Spanish. Only for taking part in this experiment you will receive 3 Euros.

2. The experiment consists of 6 periods. In each period you will be an **observer**. As an **observer** you will make **predictions** about the decisions made by participants of an experiment that was conducted in this laboratory. You will be given information about the day in which the experiment took place, the number of participants and the language chosen by the participants in that experiment as well as any other relevant information.

3. We will call the people that participated in that experiment **participants.** All **participants' decisions** were numeric and they were between 0 and 100 ECU (experimental currency unit). To understand the **decisions** made by the **participants** we will provide you with the full instructions of that experiment. Once you have read and understood the instructions you will make two types of predictions:

    a. Predictions about the **decisions** made by the **participants**.

    b. Predictions about the **predictions** made by the rest of **observers** in this room (excluding yourself).

4. The earnings that you will get at the end of this experiment will depend on how accurate your **predictions** have been. We will ask you about the average of the **decisions** and the **predictions**. Your earnings will depend on the difference between your **predictions** and those averages.

    a. If the difference is equal o below 1 ECU, you will get 5 Euros.

    b. If the difference is equal o below 5 ECU, you will get 3 Euros.

    c. If the difference is equal o below 10 ECU, you will get 1 Euro.

    d. If the difference is above 10 ECU, you will not receive anything.

5. At the end of the experiment the computer will randomly choose one period to be paid. Given that in each period you will made several predictions you can get some money. Some of your decisions will not be paid and you will be informed of that in the instructions that will appear in your computer screen. At the end of the experiment, you will get information about the period chosen to be paid, your decisions in that period, the accuracy of your predictions and your final earnings.