Foreign Language use in Competition vs. Cooperation

Florian Heine*[1], Joeri van Hugten[2], Arjen van Witteloostuijn[2], and Anne-Wil Harzing[1]

[1]Tilburg Institute of Governance, Tilburg University, PO Box 90153, 5000 LE Tilburg, Netherlands
[2]School of Business and Economics, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, Netherlands

April 2021

Abstract

We present results from an experiment using three cooperation games – prisoner's dilemma, public goods and volunteers dilemma game – conducted either in a player's mother tongue or in one of two foreign languages: English or Dutch. We find joint evidence for a mother tongue bias, such that individuals show higher cooperation levels in their mother tongue across the three games. This effect is not driven by cultural identity or perceived cooperativeness of a language's culture. While gender does seem to play a role, there is no homogeneous effect between the games.

1 Introduction

Over the last couple of decades, international trade and multinational cooperation has grown tremendously. World trade, for example, has increased more than 30 fold since 1950 (Federico & Tena-Junguito, 2017), exports have are over 4,000 times larger than in 1913 (Esteban Ortiz-Ospina & Roser, 2019). However, there is also pushback against broader cooperation, especially cross-national cooperation. For example, in May 2018, two Dutch universities and the national educational inspection were sued for offering programmes only in English. More generally, the increasing role of English as a lingua franca is subject to considerable debate, especially in the European context with the UK planning to leave the European Union. The rise of nationalist parties across the world has produced an environment where national identity, culture, and language once again take centre-stage (Kuhelj, 2011). As language is one of the most noticeable markers of (social) identity (i.e. Miller, 2000), migrants increasingly face disapproval and even violent reactions when using their native language.

These anecdotes suggest that language may play a role in this pushback against cooperation. Thus, we feel it timely to extend research on cooperation games by considering the effect of language. Specifically, do individuals make more cooperative decisions in their native language than in a foreign language? Fir this, we randomly assign German natives to play one-shot cooperation games in German, English, or Dutch.

One argument is that playing in a foreign language increases the cognitive load, which may trigger individuals to allocate more cognitive resources to the task. The availability of these resources can help finding the strategy that maximises their own personal payoff (i.e.

*Corresponding author. Tel.: +31 13 4664678. *E-mail address:* f.a.heine@tilburguniversity.edu

the Nash equilibrium). In our games, the Nash-equilibrium is no cooperation (e.g. contribute zero in the public goods game). Therefore, the foreign language may lead to less cooperation.

The concurrent argument foresees a different effect of cognitive load. Specifically, the greater cognitive load of the foreign language may leave less cognitive resources for participants to calculate the Nash equilibrium, being "no cooperation". Therefore, the foreign language may lead to more cooperation.

A third argument in this context relates to the issue of cultural identity. Being immersed in an experiment in a certain language may nudge individuals towards conforming with the (potentially subjectively) perceived degree of cooperativeness in the culture associated with the respective language. In this sense, behaviour would conform to individual beliefs about this cultural metric.

There are three prior studies on this relationship. They show a positive mother-tongue effect (i.e., individuals cooperate more in the mother tongue treatment). Akkermans, Harzing, and Van Witteloostuijn (2010) use a binary cooperation game and two languages. Gargalianou, Urbig, and Van Witteloostuijn (2017) use a binary cooperation game with one foreign language and two native languages. Urbig, Terjesen, Procher, Muehlfeld, and Van Witteloostuijn (2016) use a continuous cooperation game in a two language setting.

In this study we advance this work in two ways. *First*, we use three kinds of cooperation games. Next to a significant exploratory appeal, this feature allows to take idiosyncratic game effects into account. For example, given that we observe a common directionality among the three games, general trends in the treatment effect are likely not driven by unique (potentially unobservable) characteristics of a specific game, but generalisable features. Second, we use three languages – one native language and two foreign languages.[1] Also here, this approach allows us to reduce idiosyncratic language effects between foreign tongues. Past studies have predominantly employed English as foreign language. By introducing Dutch as second alternative foreign language, we achieve a higher degree of control over language-related specifics and can make more robust inferences.

Our results identify a common directionality among all three games. We find more cooperation in the native language consistently across the three games. The foreign language effect is strong; approximately as strong as the effect of gender on cooperation. However, the foreign-language effect also has large standard errors, possibly due to heterogeneity in the effect. We provide evidence of between-gender heterogeneity in the effect. This suggest that more research on moderators of the foreign-language effect to explore this heterogeneity may be fruitful.

2 Experimental Design

We study cooperation comparing three language treatments: *1*) Mother tongue use (German), *2*) foreign language 1 (English), and *3*) foreign language 2 (Dutch). While earlier studies have predominantly utilised English as foreign language (i.e. Akkermans et al., 2010), our study can rule out idiosyncratic results caused by specific language combinations by employing Dutch as another foreign language. While Gargalianou et al. (2017) also use a three-languages-setup, two of their languages are non-foreign languages, i.e. Dutch and French both are official languages in Belgium. We recruit German students and present a German, Dutch or English setup.

We employ three one-shot cooperation games in stranger matching and without intermediary feedback as described in the upcoming subsections. Prior to the cooperation games we take an individual measure for social value orientation (SVO), employing techniques by

[1]Each participant plays all three games in the same language treatment.

		ⓒ	
		L	R
ⓡ	T	10,10	0,18
	B	18,0	4,4

Table 1: Prisoners Dilemma Game.

Murphy, Ackermann, and Handgraaf (2011). The following subsections discuss each game and its associated equilibrium strategies in some detail, before discussing the experiment's procedures. Note that we randomise the order of the games at the treatment level. Hence, players may come across the three games in a different sequence.

2.1 Game 1 – Prisoner's Dilemma

Participants are sorted in pairs of two, with one being the row player ⓡ and the other the column player ⓒ. They simultaneously make a choice between either T and B (row player) or L and R (column player), as visualised in Table 1. This is a classical setup of the prisoner's dilemma game as described by Luce and Raiffa (1958).

Players will learn the realisation of their choices at the end of the experiment. While the only strong Nash equilibrium of this one shot game is $NE = (B, R)$, the social optimal outcome would be $SO = (T, L)$.

2.2 Game 2 – Public Goods Game

Participants are sorted in groups of four and receive an individual endowment of $T = 10$ tokens for the public goods game. Let w_i be the investment into the public good by player $i \in I$. This game employs a marginal per capita return of $MPCR = 0.5$. Individual earnings for this game are determined by:

$$\pi_i (w_i) = T - w_i + MPCR \cdot \sum_{i \in I} w_i \tag{1}$$

Results for this game will be disclosed to participants towards the end of the experiment. Under individualistic preferences the Nash equilibrium is $w_i = 0 \ \forall \ i \in I$, whereas the group's sum of individual payoffs is maximised at full contribution, i.e. $w_i = 10 \ \forall \ i \in I$.[2]

2.3 Game 3 – Volunteer's Dilemma Game

Participants are sorted in new groups of four ($N = 4$) in stranger matching. Each player $k \in K$ chooses between two options, i.e. to volunteer ($v_k = 1$) or not to volunteer ($v_k = 0$). The individual cost to volunteer is $C = 4$, which is due whenever an individual decides to volunteer – i.e. even if she is not pivotal. If at least one player volunteers, each group member receives $U = 20$, else, if nobody volunteers, they receive $D = 4$. Individual payoffs are characterised as follows:

$$\pi_k (v_k) = \begin{cases} 20 - C \cdot v_k & \text{if } \max_{k \in K} \{v_k\} > 0 \\ \\ 4 - C \cdot v_k & \text{Otherwise} \end{cases} \tag{2}$$

While most classical social dilemma games have a dominant strategy (cf. Dawes, 1974), the volunteer's dilemma formalises a type of social dilemma game in which no such dominant

[2] See formal derivation in Appendix A.1.

strategy exists (Diekmann, 1985). Clearly, any pure strategy $v_k \geq 0$ cannot be a dominant one. While $v_k = 1$ is a best response strategy to a maximin strategy of all other players of $\sum_{\substack{l \neq k \\ l \in K}} v_l = 0$, this strategy collapses if everybody adheres to this game plan. Furthermore, all players playing the maximin strategy $\sum_{k \in K} vk = 0$ constitutes no equilibrium strategy as well.

In a tacit game, absent any coordination or clear social norms, there exists a weak symmetric equilibrium point in mixed strategies, though. Let q_k be player k's probability of choosing $v_k = 0$. Then:

$$q_k = \left(\frac{C}{U}\right)^{\frac{1}{N-1}} \tag{3}$$

For the calibrations of this game, this pans out at $q_k = \left(\frac{1}{5}\right)^{\frac{1}{3}} \approx 0.585$. It can be shown that the expected earnings under this mixed strategy are no different than always contributing, hence the classification of q_k as weak.[3]

2.4 Procedures

We use ORSEE by Greiner (2004) to recruit 130 participants (mainly students of business and/or economics) for the experiment, conducted at BEElab, Maastricht University, Netherlands between October and November 2018. In this computerised laboratory experiment, participants sat in visually and physically separated cubicles, which they were randomly allocated to. The software was programmed using "z-Tree" (Fischbacher, 2007). Each participant found general instructions in the particular treatment language at her seat. Further detailed instructions about each game were displayed on the computer screen, followed by a trial round, in which participants answered understanding questions about the respective game. Before continuing towards the decision screen, all understanding questions had to be answered correctly. As memory aid, instructions-in-brief were provided on the lower part of the computer display during the decision screen.

We randomise the order of the three games at session level (all players that are connected to each other go through the game in the same order), and provide no feedback during the experiment. This means, participants only learn about their performance and the implications emanating from other players' choices, towards the end of the experiment. We run this experiment in stranger matching, meaning that between each game, the composition of groups and pairs is re-shuffled. Anonymity is guaranteed at all times.

The experiment lasted about 60 minutes including introduction, instructions, trial periods and post-experiment questionnaire. Participants earned on average € 10.01[4] with some degree of heterogeneity between the treatments (see also discussion in Appendix B).

3 Results

First we give an overview of general treatment effects (language effects) for the games, before discussing gender effects and interdependencies between the games. Prior research (i.e. Gargalianou et al., 2017) has identified heterogeneities in the language effect between male and female participants. Also in our experiment, we find some evidence for this. In Appendix B we compare general background characteristics of our participants between treatments. General demographic features and psychological metrics are well balanced between the treatments, confirming the randomisation procedures applied. By contrast, language related (like the

[3] For derivations see Appendix A.2.

[4] About US$ 11.41 at the time of the experiment.

subjective/self-assessed language ability and the foreign language ability) and culture related (for example the fit between the culture associated with the language and own personality) features display strong heterogeneity between the language treatments.

3.1 Statistical Methodology

In this study we randomise group compositions between the games and the order of the games. Furthermore, all games are set up as one-shot games without intermediate feedback – i.e. participants only learn the outcome of each individual game after the *third* game has been concluded. This randomisation allows us to analyse each individual choice per game as individual observation. Hence, we use standard OLS for tests involving interval/ratio data, including accompanying F-statistics for joined significance. For analysis on categorical/binary data, we use simple logistic regression (logit model), including χ^2 statistics for joined significance testing. For independent sample tests we use non-parametric Mann-Whitney U tests (also known as Wilcoxon rank-sum test; MWU henceforth) (Mann & Whitney, 1947) and for paired data we use the Wilcoxon matched-pairs signed-ranks test (Wilcoxon test hereafter) (Wilcoxon, 1945). For multiple hypothesis tests, involving three or more groups, we employ the Kruskal-Wallis test (KW test) (Kruskal & Wallis, 1952) and Dunn's test (Dunn, 1964) with a false discovery rate (FDR) adjustment by Benjamini and Hochberg (1995).

3.2 Overall Treatment Effects

Consider Figure 1 to see that there is joint evidence across all three games of the directionality of the treatment effect. A first glance at the data suggests a hierarchy in cooperation levels, such that they were highest in the own mother tongue, followed by Dutch, and lowest when the games were played in English language.

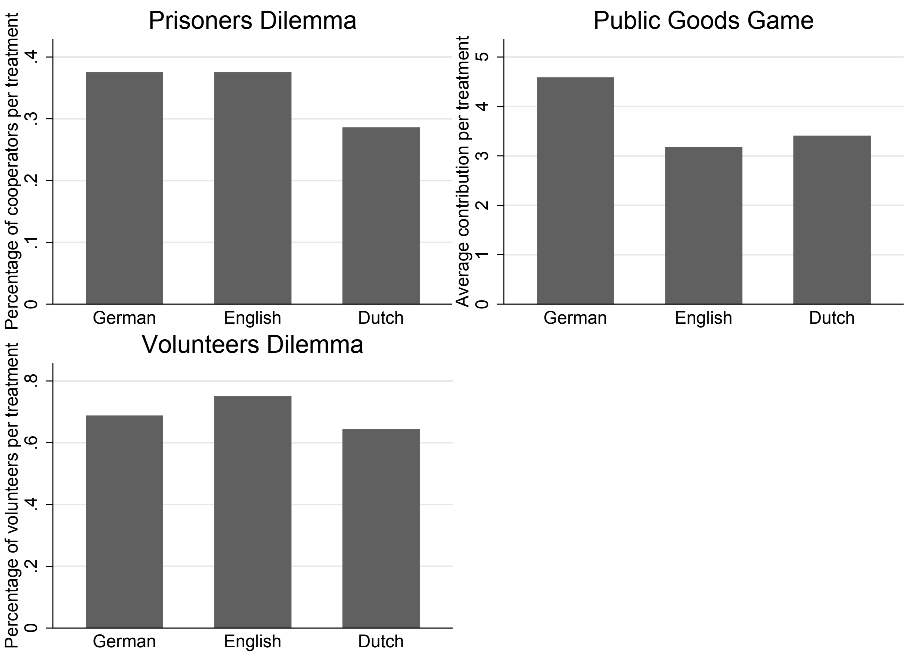

Figure 1: Comparing Cooperation Levels between our Games and the Languages

The following subsections investigate each treatment separately.

3.3 Prisoner's Dilemma

Table 2 provides the outputs for logit regressions investigating individual behaviour in the prisoner's dilemma game. We observe no significant treatment difference in this language treatment. Only two factors appear to have significant descriptive power: The decision time players spend on this game negatively influences their willingness to cooperate. The decision time includes the total time spent in this game, including reading of instructions, the trial round and the actual strategic decision. In Subsection 3.8 we show that players spend a few seconds more on this game in the Dutch language treatment.

Furthermore, *perceived cooperativeness*[5] has a strong positive effect on cooperation levels. This suggests that in this game, participants mimic the perceived level of cooperativeness in their treatment language.

VARIABLES	(1) Cooperative Play	(2)	(3)	(4)
English	0.000	0.042	0.046	−0.269
	(0.44)	(0.45)	(0.46)	(0.67)
Dutch	−0.405	−0.199	−0.092	0.290
	(0.45)	(0.47)	(0.50)	(0.68)
Female		0.174	0.024	0.133
		(0.39)	(0.40)	(0.42)
SVO angle		0.018	0.019	0.023
		(0.01)	(0.01)	(0.02)
Culture fit			−0.139	−0.134
			(0.14)	(0.15)
Dec. Time Prisoners Dilemma Game			−0.012**	−0.013**
			(0.01)	(0.01)
Perceived Cooperativeness				1.036***
				(0.39)
Creative				−0.091
				(0.13)
Structured				−0.075
				(0.16)
Restricted				−0.187
				(0.13)
Complex				−0.125
				(0.13)
Constant	−0.511*	−1.158**	1.318	−1.639
	(0.30)	(0.54)	(1.23)	(2.51)
N	130	127	127	127
Pseudo R-squared	0.006	0.016	0.053	0.119

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

Standard errors in parentheses.

Table 2: Logit Regression: Is there a difference between treatments with respect to cooperative play in the prisoner's dilemma game?

[5] The perception of a language's cooperativeness is constructed by the average of *equality*, *fair*, *individualistic*, *masculine*, *just* and *egoistic*, as in Akkermans et al. (2010); House, Hanges, Javidan, Dorfman, and Gupta (2004). The factors *equality-egoistic* are described in Appendix B.

3.4 Public Goods Game

Table 3 provides the outputs of OLS regressions for the public goods game. Regressions (5)-(7) indicate that participants show a lower contribution level in the English treatment than those in the German one. Also in the Dutch treatment, a clear trend is visible towards a lower contribution level, when compared to the German treatment. When testing against the other two language treatments, participants in the German treatment contribute significantly more than in the two other groups.[6] This finding is confirmed by a KW test[7] and Dunn's test (Table 4). In specific, contribution to the group account is higher in the mother tongue, German, than in any of the two other languages. When comparing contribution behaviour in the English and Dutch language treatment, we can find no significant differences.

	(5)	(6)	(7)	(8)
VARIABLES	Contribution to Group Account			
English	-1.408^{*}	-1.338^{*}	-1.272^{*}	-2.590^{***}
	(0.72)	(0.69)	(0.70)	(0.92)
Dutch	-1.179^{*}	-0.854	-0.956	-2.693^{***}
	(0.71)	(0.70)	(0.72)	(0.93)
Female		0.708	0.645	0.608
		(0.58)	(0.59)	(0.59)
SVO angle		0.057^{***}	0.057^{***}	0.059^{***}
		(0.02)	(0.02)	(0.02)
Culture fit			-0.035	-0.039
			(0.21)	(0.21)
Dec. Time Public Goods Game			0.001	0.001
			(0.00)	(0.00)
Perceived Cooperativeness				-0.363
				(0.51)
Creative				0.297^{*}
				(0.18)
Structured				-0.576^{***}
				(0.21)
Restricted				0.158
				(0.18)
Complex				-0.182
				(0.17)
Constant	4.583^{***}	2.567^{***}	2.379	7.095^{**}
	(0.48)	(0.78)	(1.55)	(3.26)
N	130	127	127	127
R-squared	0.035	0.102	0.105	0.181

* p<0.10, ** p<0.05, *** p<0.01
Standard errors in parentheses.

Table 3: OLS Regression: Is there a difference between treatments with respect to contributions to the group account in the public goods game?

[6] MWU test: H0: diff = 0, H1: diff $\neq$ 0, N = 130. P = 0.015. Higher rank sum than expected for German treatment.

[7] KW test with ties: N = 130, P = 0.052

Column mean - row mean z test statistic (*p*-value)	German	English
English	3.67***	
	(0.000)	
Dutch	3.55***	−0.16
	(0.000)	(0.437)

* p<0.10, ** p<0.05, *** p<0.01
p-values in parentheses

Table 4: Dunn's Pairwise Comparison of contribution by language treatment (Benjamini-Hochberg)

Furthermore, individuals who display a more prosocial orientation through the *SVO angle* also contribute more to the public goods game. Players who perceive the language they are in as more *creative* tend to contribute more, while those who perceive their treatment language as more *structured* tend to contribute less.

To make the analysis of the three games more comparable – and to test if differences in power between the games are a mere result from the lower power in statistical analysis of binary data, or if individuals react differently to binary games, than if confronted with a more nuanced choice set – we convert the data from the public goods game into binary data. 0 if the individual contributes nothing, 1 if she makes a contribution larger than zero.

Table 5 shows the regression outputs for logit regressions of

3.5 Volunteers Dilemma

In the volunteers dilemma game, individual choice cannot be explained by the treatment language or other factors we included in our hypotheses. Accordingly, the regression outputs from Table 6 show very high *p*-values overall and a very small pseudo R-squared. Players' behaviour in this game cannot be explained by any of the factors we collected. The binary nature of this game, along with its single-shot character, makes it unconducive for this kind of analysis.

3.6 Gender Differences

As in (Gargalianou et al., 2017), also in our study there exist gender differences between the language treatments. Again, this is only significant for the public goods game. Both Table 7 and Figure 2 show that in the public goods game, the German language treatment is the only one in which contribution levels are on a similar level both for participants identifying as male and female. In the two other language treatments, by contrast, either the male or the female counterparts display a significantly lower contribution level, which might contribute at explaining the treatment differences we observe for this game. Regression (14) suggests that treatment differences in contribution to the public goods game be driven by either male participants in the English treatment or female participants in the Dutch treatment.[8]

[8] Our randomisation checks in Appendix B do not lend evidence for this effect to be driven by issues with the randomisation in our study.

VARIABLES	(9)	(10)	(11)	(12)
	Contribution to Group Account			
English	−0.640	−0.724	−0.741	−2.346***
	(0.52)	(0.58)	(0.58)	(0.87)
Dutch	−0.916*	−0.648	−0.629	−1.970**
	(0.51)	(0.56)	(0.57)	(0.82)
Female		1.170**	1.170**	1.343***
		(0.46)	(0.47)	(0.50)
SVO angle		0.055***	0.054***	0.061***
		(0.02)	(0.02)	(0.02)
Culture fit			0.038	0.030
			(0.16)	(0.17)
Dec. Time Prisoners Dilemma Game			−0.000	0.000
			(0.01)	(0.01)
Perceived Cooperativeness				−0.268
				(0.43)
Creative				0.073
				(0.15)
Structured				−0.295
				(0.19)
Restricted				−0.150
				(0.16)
Complex				−0.362**
				(0.16)
Constant	1.609***	−0.284	−0.452	4.627
	(0.39)	(0.57)	(1.42)	(2.99)
N	130	127	127	127
Pseudo R-squared	0.024	0.159	0.160	0.222

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

Standard errors in parentheses.

Table 5: Logit Regression using binary data from the public goods game. 0 if contribution is zero, 1 otherwise.

3.7 Cooperativeness Between Games

In this subsection we investigate the interdependence between the three games. We employ both the Pearson correlation coefficient (including for binary games the point biserial correlation coefficient) and Spearman's rank correlation (Spearman, 1987) with substantially equivalent results. Table 8 represents the Pearson correlation coefficient matrix for all pairwise correlations between the games. Two of the three possible pairwise correlations are significantly positive, albeit of only moderate size. The only pairwise correlation that is not significantly different from zero is between the public goods game and the volunteers dilemma.

3.8 Decision Time

Consider Figure 3 to see that there appears to be a slight tendency for participants to spend more time in the Dutch language treatment, than in the other ones. For each game, the decision time includes the time a participant is on an active screen during the instructions,

VARIABLES	(13) Cooperative Play	(14)	(15)	(16)
English	0.310	0.328	0.365	0.435
	(0.48)	(0.48)	(0.49)	(0.68)
Dutch	−0.201	−0.106	−0.387	−0.789
	(0.45)	(0.46)	(0.53)	(0.73)
Female		0.013	−0.010	−0.108
		(0.40)	(0.40)	(0.41)
SVO angle		0.010	0.011	0.006
		(0.01)	(0.01)	(0.01)
Culture fit			−0.045	−0.090
			(0.14)	(0.15)
Dec. Time Volunteers			0.003	0.003
Dilemma Game			(0.00)	(0.00)
Perceived				−0.593
Cooperativeness				(0.37)
Creative				0.163
				(0.13)
Structured				0.029
				(0.15)
Restricted				0.018
				(0.13)
Complex				0.070
				(0.13)
Constant	0.788**	0.493	0.331	2.276
	(0.31)	(0.52)	(1.04)	(2.26)
N	130	127	127	127
Pseudo R-squared	0.007	0.010	0.018	0.047

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

Standard errors in parentheses.

Table 6: Logit Regression: Is there a difference between treatments with respect to cooperative play in the volunteers dilemma game?

the trial period and the actual decision of a game. It does not include time a participant waits for other players. Note further, that *Total Decision Times* includes more than only the sum of the three games. For this measure we also include time spent in other transitioning screens, the final payment overview and instructions for the SVO measure.

Regressions in Table 9 confirm that indeed participants spend more time on their decisions in the Dutch language treatment. This is true for all games and the total time spent in the experiment. Concerning the English language treatment, participants seem to spend about the same time as those in the mother tongue setting do.

Decision time could be a proxy for fluency in a language, such that participants require more time to understand the instructions and the implications of their choices in a language setting they are less fluid in. In the same way, considering that all three games were played as one-shot and one-decision games, decision time may reflect the (perceived) complexity of a game. In this sense, participants take significantly more time for the public goods game

VARIABLES	(17) Prisoners Dilemma Cooperative Play in each game	(18) Public Goods Game	(19) Volunteers Dilemma
English	0.087	−3.298***	0.492
	(0.76)	(1.09)	(0.77)
Dutch	0.182	−0.235	0.087
	(0.71)	(1.02)	(0.69)
Female	0.470	0.365	0.241
	(0.65)	(0.95)	(0.65)
English × Female	−0.018	3.364**	−0.241
	(0.94)	(1.39)	(0.99)
Dutch × Female	−0.806	−1.786	−0.395
	(0.95)	(1.37)	(0.93)
Constant	−0.875	4.235***	0.606
	(0.53)	(0.76)	(0.51)
N	127	127	127
(Pseudo) R-squared	0.011	0.144	0.008

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

Standard errors in parentheses.

Table 7: Logit/OLS Regressions investigating the gender effect, including interaction terms between female and treatment number.

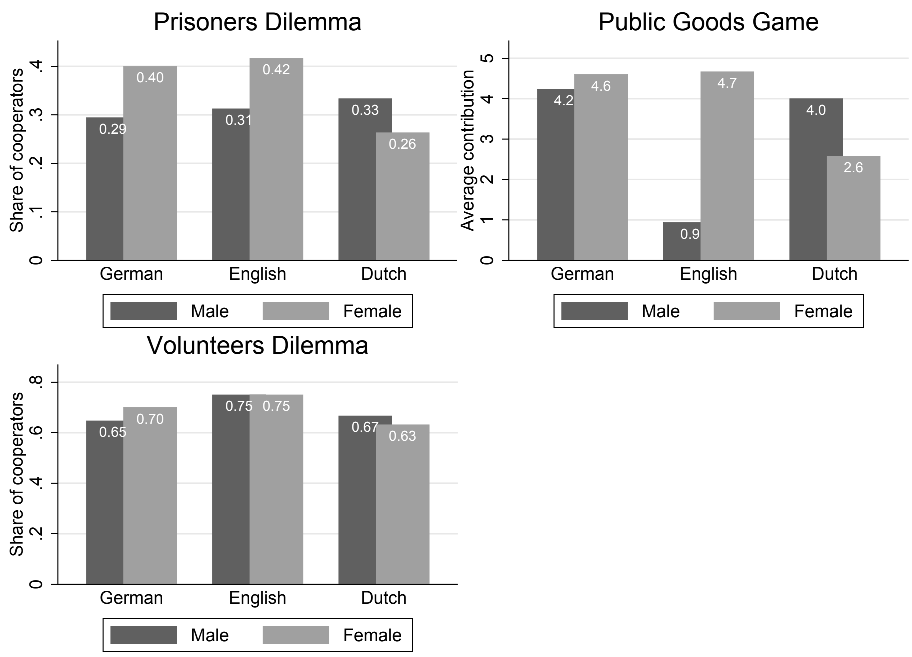

Figure 2: Comparing Cooperation Levels between players identifying as male or female across the games and the language treatments

than for either the prisoners dilemma[9] or the volunteers dilemma games,[10] respectively.

[9] Wilcoxon test : $N = 130 \quad p = 0.000$

[10] Wilcoxon test : $N = 130 \quad p = 0.000$

correlation coefficient (p-value)	Prisoners Dilemma	Public Goods Game	Volunteers Dilemma
Prisoners Dilemma	1.000		
Public Goods Game	0.146* (0.098)	1.000	
Volunteers Dilemma	0.24*** (0.006)	−0.011 (0.901)	1.000

* p<0.10, ** p<0.05, *** p<0.01
p-values in parentheses.

Table 8: Pearson's pairwise correlation coefficient between the language treatments

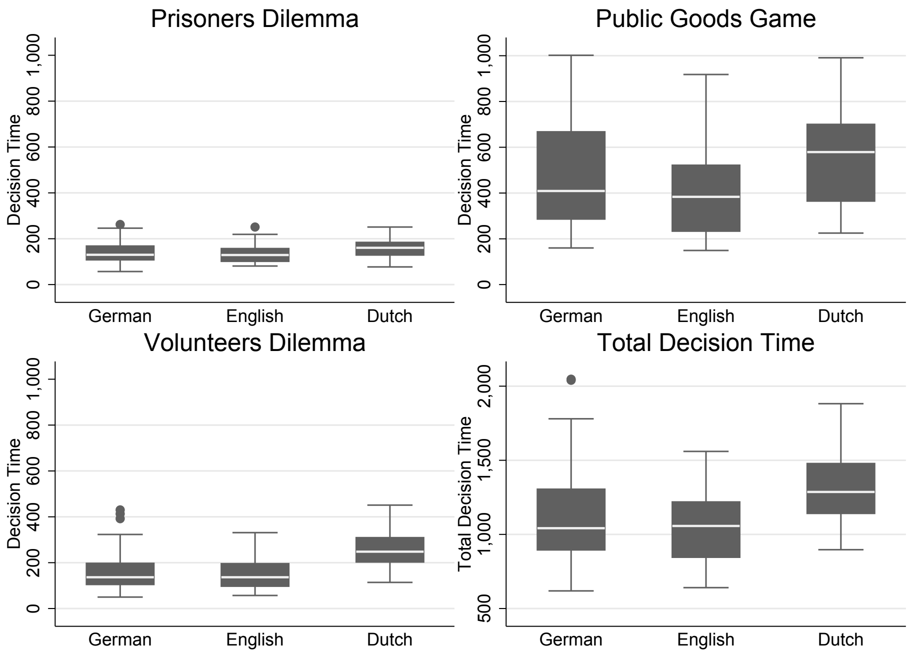

Figure 3: How much time do participants spend in each game per language treatment. Note that Total Decision Time uses a different y axis scale.

4 Conclusion

In this article we compare pro-social behaviour in three cooperation/coordination games between three language conditions. We conduct an experiment using German participants in either German (mother tongue), English or Dutch. There exists a tendency towards common directionality among all three games, such that players show highest cooperation levels in their mother tongue. Differences between the two foreign language conditions, Dutch and English, are only marginal in all treatments and are far from being significant, even under the presence of control factors.

As earlier studies have showcased as well (i.e. Gargalianou et al., 2017), parts of our results seem to be driven by gender effects. This however, does not show a common directionality, such that male participants seem to cooperate somewhat less in the English language treatment, while female participants have a slightly lower cooperation level in the Dutch language treatment.

VARIABLES	(20) Total Time	(21) Time Prisoner's Dilemma	(22) Time Public Goods Game	(23) Time Volunteers Dilemma
English	−72.658	−2.750	−68.429	-12.363
	(59.48)	(8.64)	(48.04)	(17.57)
Dutch	202.530***	18.107**	84.926*	89.289***
	(58.71)	(8.52)	(47.42)	(17.34)
Constant	1125.208***	137.250***	478.979***	164.688***
	(40.10)	(5.82)	(32.39)	(11.84)
N	130	130	130	130
R-squared	0.147	0.050	0.070	0.233

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

Standard errors in parentheses.

Table 9: OLS Regressions investigating differences in decision times between the language treatments.

Our data shows that participants needed significantly more time for the public goods game, and in the Dutch language treatment. We interpret decision time as a proxy for complexity on two dimensions: *1*) Participants appear to require more time for games that appear to be more complex. In this regard, the public goods game seems to require the highest degree of cognitive capacity, among the three games. At the same time, the non-binary, graded character of this game provides the highest degree of power in the analysis. For future research, we regard this feature to be of pivotal importance and probably worthwhile the extra effort towards conveying the (apparently more difficult) instructions to the participants. *2*) Participants require more time in a language that they perceive as more difficult. Our results demonstrate that this aspect does not seem to influence cooperation levels, i.e. we do not observe strong differences in cooperation levels between the two foreign language treatments, albeit stark differences in terms of decision time.

Appendix A Risk Neutral Equilibria

A.1 Game 2 – Public Goods Game

Players individually maximise their profit π_i by setting own contribution w_i:

$$\pi_i\left(w_i\right) = T - w_i + MPCR \cdot \sum_{i \in I} w_i \tag{4}$$

with $MPCR = 0.5$. As $\pi_i' < 0$ and $\pi_i'' = 0$, there exists a corner solution $w_i = 0$.

A.2 Game 3 – Volunteer's Dilemma Game

This analysis closely follows techniques described in Diekmann (1985), variables are defined as in Subsection 2.3. Expected individual payoff for a player k in the mixed strategy setting is characterised as:

$$E_k\left(q_k\right) = q_k \cdot U \left(1 - \pi^{N}_{\substack{j \neq k \\ j \in K}} q_j\right) + \left(1 - q_k\right) \cdot \left(U - C\right) \tag{5}$$

After deriving with respect to q_k we arrive at the best response function:

$$\frac{\partial E_k}{\partial q_k} = 0 \quad \Leftrightarrow \quad C - U\pi^{N}_{\substack{j \neq k \\ j \in K}} q_j = 0 \tag{6}$$

This solves for $q_k^O = \left(\frac{C}{U}\right)^{\frac{1}{N-1}}$. Plugging this into Equation 5:

$$E_k\left(q_k^O\right) = \left(\frac{C}{U}\right)^{\frac{1}{N-1}} \cdot U\left(1 - \left(\frac{C}{U}\right)^{\frac{N-1}{N-1}}\right) + \left(1 - \left(\frac{C}{U}\right)^{\frac{1}{N-1}}\right) \cdot (U - C) \tag{7}$$

Equation 7 solves for $E_k\left(q_k^O\right) = U - C$, which is equal to the maximin strategy of always contributing.

Appendix B Randomisation Check

Table 10 depicts general background characteristics of the experiment participants. For general demographics and general psychometric measures, χ^2 or F-statistics show no systematic difference between the treatments.

On what concerns culture related factors, we do observe significant differences between the language treatments. This is a set of culture related aspects, participants rate at the start of the experiment on a 7-point Likert scale. For the factor *culture fit* participants indicate, how well the culture they associate with the language they are in fits with themselves. For the other *culture related* items, participants mark the point between two contrasting features (i.e. inequality/equality, creative/uncreative), indicating their assessment of proximity towards the two extremes of the language of their treatment. For this, the factor "Complex" (opposed to "Simple") stands out enormously. Participants perceive the culture they associate with the German language as a much more complex than the one associated with the Dutch or English language, respectively. Another interesting hierarchy is "Structured" (opposed to "Unstructured"), where participants recognise a higher degree of structure in the culture they associate with the German language, than English or Dutch.

We observe a similarly heterogeneous effect for language related general metrics like the subjective language ability (SLA)[11] or the Foreign Language Anxiety (FLA)[12]. In this regard, individuals rate their language abilities lowest in Dutch[13] and have the highest mean FLA in this language.[14] Most curiously though, with regards to subjective language ability, English ranks more favourably than their mother tongue, German.[15] This might be related to the fact that the language of instruction at the faculty where the experiment was conducted is almost exclusively English, so students feel comfortable using this language.

Furthermore, there exists some heterogeneity concerning the amount of earnings between the treatment languages. In particular, participants in the English treatment earned significantly more than their counterparts from the other treatments.[16]

[11] 8 item Likert scale asking 'How would you assess your ability to understand written English?' or a translation of the question enquiring about the SLA of the current language.

[12] 5 vignettes of 7 item Likert scales. For example: 'My heart starts pounding when I'm going to be called on in a meeting in English more than in a meeting in my native language' or a translation of the question.

[13] MWU test: H0: diff = 0, H1: diff $\neq$ 0, N = 76. P = 0.000. Lower rank sum than expected for Dutch treatment.

[14] MWU test: H0: diff = 0, H1: diff $\neq$ 0, N = 130. P = 0.000. Higher rank sum than expected for Dutch treatment.

[15] MWU test subjective language ability: H0: diff = 0, H1: diff $\neq$ 0, N = 64. P = 0.000. Lower rank sum than expected for German treatment.

[16] MWU test: H0: diff = 0, H1: diff $\neq$ 0, N = 130. P = 0.000. Higher rank sum than expected for English treatment.

Treatment	German ($n = 48$)	English ($n = 40$)	Dutch ($n = 42$)	χ^2 or F
Demographics				
Age (mean years)	20.96	21.02	21.58	0.3
Female	0.64	0.60	0.47	2.51
Psychological Metrics				
SVO angle	25.52	26.90	23.15	0.79
Efficiency orientation	1.40	1.31	1.50	0.86
Avg. distance to Equality	0.25	0.25	0.24	0.03
Avg. distance to Altruism	0.70	0.69	0.69	0.02
Culture Related				
Culture fit	5.79	6.02	5.19	3.74**
Equality	5.81	6.22	5.86	0.93
Creative	3.77	4.43	4.75	3.73**
Fair	2.9	3.42	2.56	3.19**
Structured	6.13	5.31	4.94	7.96***
Restricted	2.94	2.68	3.61	3.5**
Individualistic	4.64	3.79	4.03	2.54*
Complex	5.3	2.22	3	40.09***
Masculine	4.08	3.97	3.39	3.47**
Just	5.79	5.32	5.47	1.17
Egoistic	4.57	4.38	3.42	7.97***
Language Related				
Subjective language ability	5.50	6.90	2.67	78.52***
Foreign Language Anxiety (mean)	3.41	3.07	4.05	5.81***
Other				
Earnings in points	47.28	59.51	44.27	21.07***

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

χ^2 statistic reported for difference in categorical variable between treatments;

F-statistic for difference in continuous variable.

Table 10: Individual Background Characteristics

References

Akkermans, D., Harzing, A.-W., & Van Witteloostuijn, A. (2010). Cultural accommodation and language priming. *Management International Review*, *50*(5), 559–583.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300.

Dawes, R. M. (1974). *Formal models of dilemmas in social decision-making* (Tech. Rep.). OREGON RESEARCH INST EUGENE.

Diekmann, A. (1985). Volunteer's dilemma. *Journal of Conflict Resolution*, *29*(4), 605-610. Retrieved from https://doi.org/10.1177/0022002785029004003 doi: 10.1177/0022002785029004003

Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics*, *6*(3), 241–252.

Esteban Ortiz-Ospina, D. B., & Roser, M. (2019). Trade and globalization. *Our World in*

Data. (https://ourworldindata.org/trade-and-globalization)

Federico, G., & Tena-Junguito, A. (2017, Aug 01). A tale of two globalizations: gains from trade and openness 1800–2010. *Review of World Economics*, *153*(3), 601–626. Retrieved from https://doi.org/10.1007/s10290-017-0279-z doi: 10.1007/s10290-017-0279-z

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171-178.

Gargalianou, V., Urbig, D., & Van Witteloostuijn, A. (2017). Cooperating or competing in three languages: cultural accommodation or alienation? *Cross Cultural & Strategic Management*, *24*(1), 167–191.

Greiner, B. (2004). The online recruitment system orsee 2.0. *A guide for the organization of experiments in economics. University of Cologne*.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (2004). *Culture, leadership, and organizations: The globe study of 62 societies*. Sage publications.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, *47*(260), 583–621.

Kuhelj, A. (2011). Rise of xenophobic nationalism in europe: A case of slovenia. *Communist and Post-Communist Studies*, *44*(4), 271 - 282. Retrieved from http://www.sciencedirect.com/science/article/pii/S0967067X11000468 doi: https://doi.org/10.1016/j.postcomstud.2011.10.003

Luce, R. D., & Raiffa, H. (1958). *Games and decisions: Introduction and critical survey: a study of the behavioral models project...* J. Wiley & sons.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, *18*(1), 50-60.

Miller, J. M. (2000). Language use, identity, and social interaction: Migrant students in australia. *Research on Language and Social Interaction*, *33*(1), 69-100. Retrieved from https://doi.org/10.1207/S15327973RLSI3301_3 doi: 10.1207/S15327973RLSI3301_3

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. J. (2011). Measuring social value orientation. *Judgment and Decision Making*, *6*(8), 771–781.

Spearman, C. (1987). The proof and measurement of association between two things. *The American journal of psychology*, *100*(3/4), 441–471.

Urbig, D., Terjesen, S., Procher, V., Muehlfeld, K., & Van Witteloostuijn, A. (2016). Come on and take a free ride: Contributing to public goods in native and foreign language settings. *Academy of Management Learning & Education*, *15*(2), 268-286. Retrieved from https://doi.org/10.5465/amle.2014.0338 doi: 10.5465/amle.2014.0338

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, *1*(6), pp. 80-83. Retrieved from http://www.jstor.org/stable/3001968