CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

Discussion Paper No. 2021-03

Till O. Weber, Benjamin Beranek, Simon Gächter, Fatima Lambarraa-Lehnhardt, and Jonathan F. Schulz

June 2021

The Behavioural Mechanisms of Voluntary Cooperation in WEIRD and Non-WEIRD Societies

CeDEx Discussion Paper Series
ISSN 1749 - 3293

CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

The Centre for Decision Research and Experimental Economics was founded in 2000, and is based in the School of Economics at the University of Nottingham.

The focus for the Centre is research into individual and strategic decision-making using a combination of theoretical and experimental methods. On the theory side, members of the Centre investigate individual choice under uncertainty, cooperative and non-cooperative game theory, as well as theories of psychology, bounded rationality and evolutionary game theory. Members of the Centre have applied experimental methods in the fields of public economics, individual choice under risk and uncertainty, strategic interaction, and the performance of auctions, markets and other economic institutions. Much of the Centre's research involves collaborative projects with researchers from other departments in the UK and overseas.

Please visit http://www.nottingham.ac.uk/cedex for more information about the Centre or contact

Suzanne Robey
Centre for Decision Research and Experimental Economics
School of Economics
University of Nottingham
University Park
Nottingham
NG7 2RD
Tel: +44 (0)115 95 14763
suzanne.robey@nottingham.ac.uk

The full list of CeDEx Discussion Papers is available at

http://www.nottingham.ac.uk/cedex/publications/discussion-papers/index.aspx

The Behavioural Mechanisms of Voluntary Cooperation in WEIRD and Non-WEIRD Societies*

Till O. Weber[1,†], Benjamin Beranek[2], Simon Gächter[3,4,5] ,
Fatima Lambarraa-Lehnhardt[6,7], and Jonathan F. Schulz[8]

[1] Newcastle University Business School, Newcastle University, UK; tilloweber@gmail.com
[2] Department of Economics, Missouri State University, USA; benjaminberanek@missouristate.edu
[3] School of Economics, University of Nottingham, UK; simon.gaechter@nottingham.ac.uk
[4] IZA Bonn, Germany; [5] CESifo Munich, Germany
[6] Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg, Germany;
fatima.lehnhardt@zalf.de [7] CRC-PEG, University of Göttingen, Germany
[8] Department of Economics, George Mason University, USA, jonathan.schulz77@gmail.com
[†] Corresponding author

June 12, 2021

Abstract

We provide a framework to uncover behavioural mechanisms driving potential cross-societal differences in voluntary cooperation. We deploy our framework in one-shot public goods experiments in the US and the UK, and in Morocco and Turkey. We find that cooperation is higher in the US and UK than in Morocco and Turkey. Our framework shows that this result is driven mostly by differences in beliefs rather than in cooperative preferences, or peer punishment, which are both similar in the four subject pools. Our results highlight the central role of beliefs in explaining differences in voluntary cooperation within and across societies.

Keywords: voluntary cooperation; experiments; public goods; cross-societal differences; behavioural framework.

(JEL: C9, H4, C7, D2)

*Acknowledgements: We thank Stefania Bortoletti, Lucas Molleman, Christian Thöni, Ori Weisel and participants of the Royal Economic Society Conference 2018, the 2019 Stanford Conference on Behavioural Economics and Development and various workshops for their valuable comments. This work was supported by the European Research Council [Grant Number ERC-AdG 295707 COOPERATION] and the Economic and Social Research Council [Grant Number ES/K002201/1].

1. Introduction

Trust and social capital are crucial for economic prosperity, but vary substantially across societies (e.g., Knack and Keefer (1997); Guiso et al. (2008); Tabellini (2008); Fehr (2009); Algan and Cahuc (2013)). Social capital is often expressed as voluntary cooperation in situations where collective welfare and self-interest are in conflict. Here, we apply a cross-societal perspective and use laboratory public goods experiments to uncover the behavioural mechanisms that underpin the variation in voluntary cooperation.

One important insight from the last couple of decades of behavioural economics research into voluntary cooperation is that many people have a *conditionally cooperative disposition*: they are willing to cooperate, provided others (are believed to) cooperate as well.[1] An important implication is that *beliefs* matter: the higher beliefs about others' cooperation are, the higher is own cooperation (e.g., Fischbacher and Gächter (2010); Gächter and Renner (2018)). A further robust result is that people are willing to *punish* group members even if this is costly for the punisher (e.g., Fehr and Gächter (2002); Herrmann et al. (2008); Carpenter and Matthews (2012); Weber et al. (2018)).[2]

With a few exceptions, which we review in Section 2, most results on the behavioural mechanisms of conditional cooperation, beliefs, and punishment come from the US and Western Europe – that is, countries that anthropologists and cultural psychologists Henrich et al. (2010) have labelled WEIRD: Western, Educated, Industrialized, Rich and Democratic (see also Henrich (2020) and Apicella et al. (2020)). In this paper, we test the generalizability of the behavioural mechanisms that underpin voluntary cooperation in an analysis where we compare conditionally cooperative dispositions, beliefs, and punishment in two classically WEIRD societies – the US and the UK – and two societies – Morocco and Turkey – that are culturally among the most distant countries from the US and the UK in a quantifiable way (Muthukrishna et al. (2020), see details in Section 3).

Our goal is to go beyond simply observing whether there are differences in voluntary cooperation levels in different societies (there are, see, e.g., Herrmann et al. (2008)). Instead, we provide precise measurements of the two behavioural mechanisms that arguably underpin voluntary cooperation. The first behavioural mechanism comprises preferences for conditional cooperation and beliefs about others' cooperativeness, which jointly determine cooperation in

[1] In lab experiments, e.g., Keser and van Winden (2000); Fischbacher et al. (2001); Gächter et al. (2017); in lab in the field experiments, e.g., Rustagi et al. (2010); in experiments with representative samples, e.g., Thöni et al. (2012); Fosgaard et al. (2014); in field experiments, e.g., Frey and Meier (2004). See also Gächter (2007) and Thöni and Volk (2018) for reviews.

[2] For reviews see Gächter and Herrmann (2009); Chaudhuri (2011); Fehr and Schurtenberger (2018).

the absence of punishment. The second behavioural mechanism consists of a willingness to punish free riders.

Our approach, detailed in Section 3, opens the “black box” of voluntary cooperation because it allows us to pinpoint where in the behavioural mechanisms – if at all – cross-societal differences in voluntary cooperation occur. Our paper thus does not aim to investigate how historical, cultural, or institutional factors causally impact voluntary cooperation. Its goal is to contribute to the literature that aims to better understand the mechanisms of how cultural differences might influence economic outcomes (e.g., Guiso et al. (2006); Alesina and Giuliano (2015)). This literature emphasizes the role of both beliefs and preferences as distinct aspects of culture (see e.g. Greif 1994 on the role of cultural beliefs), yet, most empirical papers do not or are unable to distinguish between beliefs and preferences (see Alesina and Giuliano (2015) for a discussion). By eliciting both – beliefs and preferences (i.e., cooperative dispositions) – we shed light on their relative importance in accounting for societal differences in voluntary cooperation among countries that vary substantially in their culture.

We use anonymous laboratory one-shot public goods games that allow us to measure the elements of the two behavioural mechanisms without having to deal with potential confounds coming from repeated play. Both mechanisms consist of a sequence of incentivized experiments, which all participants play in the same order. To test the first behavioural mechanism, participants play a one-shot public goods game in which they make two types of decisions: (*i*) an unconditional contribution to the public good, and (*ii*) contribution decisions that are conditional on other group members’ average contributions. The conditional decisions allow us to determine a participant’s *cooperative disposition.* We also elicit *beliefs about others’ cooperativeness*. We predict that preferences and beliefs jointly determine cooperation.

For the second behavioural mechanism, participants play a one-shot public goods game with peer punishment. We focus on preferences for *enforcement of cooperation* by studying how each participant’s expected punishment affects their contribution decision. The one-shot design removes any strategic incentives to engage in punishment as well as the possibility of feuds and spill-over effects over rounds (e.g., Nikiforakis (2008); Nikiforakis and Engelmann (2011)). In these non-strategic settings, any willingness to cooperate or pay for punishment cannot be motivated by strategic concerns and is therefore reflecting a “preference”.

Our main results are as follows. Overall, the majority of people have conditionally cooperative dispositions in all subject pools, and differences between them are small. The same holds for antisocial punishment. The main difference between subject pools is in the beliefs: In both experiments, UK and US subjects are more optimistic about others’ contributions and therefore also contribute more to the public good than the Moroccan and Turkish. If punishment is available, UK and US subjects are more likely to expect punishment of free riding.

We conclude by noting that the applicability of our framework presented in this paper extends beyond the cross-societal context in which we apply it here. Our framework is a "toolbox" that can be used to study voluntary cooperation of any group of people, with or without the goal of comparing groups.

2. Evidence on WEIRD and non-WEIRD voluntary cooperation

Our paper builds on, and contributes to, a literature in behavioural economics that has documented some profound cross-societal differences in social preferences as observed in economic experiments (see, e.g., Gächter et al. (2010); Thöni (2019) for overviews).[3] In this section, we review the literature most closely related to the behavioural mechanisms of voluntary cooperation. We are particularly interested in how existing societal differences in voluntary cooperation can be accounted for by the behavioural mechanism of conditionally cooperative dispositions and beliefs, which together can explain actual cooperation levels in one-shot games conducted in WEIRD subject pools (e.g., Fischbacher et al. (2012); Gächter et al. (2017)). We also review the literature on the second behavioural mechanism necessary to sustain voluntary cooperation: punishment of free riding.

An important element of our first behavioural mechanism is *conditional cooperation*. One empirical observation on how societal conditions might affect conditional cooperation is via betrayal aversion (the aversion of being let down by another person as compared to bad luck). Betrayal aversion is higher among conditional cooperators (Cubitt et al. (2017)) and it has been shown to vary between societies (Bohnet et al. (2008); Bohnet et al. (2010)). If betrayal aversion is higher among conditional cooperators and more prevalent in some societies compared to others, then the proportion of conditional cooperators might vary across societies. Another potential channel of cross-societal variability in conditional cooperation is via positive reciprocity because conditional cooperation is arguably an instance of positive reciprocity.

[3] For instance, Henrich et al. (2005) documented differences in ultimatum and dictator game behaviour (in small-scale societies), see also Oosterbeek et al. (2004). Bohnet et al. (2008) and Bohnet et al. (2010) showed cross-societal differences in betrayal aversion; Chuah et al. (2016) observed cross-societal differences in trust, religiosity and discrimination; Romano et al. (2017) investigated ingroup and outgroup trust across 17 societies; Molleman and Gächter (2018) studied cross-societal differences in social learning and cooperation; and Gächter and Schulz (2016) and Cohn et al. (2019) found that honesty differs across societies. The biggest effort in establishing broad knowledge about how economic preferences (time, risk, and social) are distributed around the world is the Global Preference Survey (Falk et al. (2018)). For a seminal paper on cross-societal differences using observational data (on norms of corruption) see Fisman and Miguel (2007).

According to evidence from the Global Preference Survey (Falk et al. (2018)), positive reciprocity varies substantially across societies.

While betrayal aversion and positive reciprocity are potential arguments why preferences for conditional cooperation might differ between societies, the existing evidence suggests that differences between countries are small. However, the large majority of experiments that have elicited conditionally cooperative dispositions have been conducted in the US, UK and other WEIRD societies (Thöni and Volk (2018))[4]. Exceptions include Herrmann and Thöni (2009); Martinsson et al. (2013) and Rustagi et al. (2010). Herrmann and Thöni (2009) conducted experiments in four Russian subject pools finding little differences in conditional cooperation between them and largely comparable results to subject pools in Switzerland (Fischbacher et al. (2001)). Martinsson et al. (2013) elicited conditionally cooperative dispositions in Vietnam and Colombia, again with similar results to those from WEIRD subject pools. Finally, Rustagi et al. (2010) elicited preferences for conditional cooperation in community forest groups in Ethiopia and found similar rates of conditional cooperation compared to those reported from WEIRD societies (Thöni and Volk (2018)). The Rustagi et al. study also shows that communities with a higher share of conditional cooperators achieve better outcomes in a naturally occurring public good – forest quality.[5]

Beliefs about others' cooperativeness are a further element of our first behavioural mechanism. Beliefs have not been regularly studied in a cross-societal context. Yet, there are several pieces of evidence that lead us to expect higher cooperative beliefs in WEIRD compared to non-WEIRD societies. For instance, World Values survey data (Inglehart et al. (2014)) suggest that the share of people who believe that "most people are trying to take advantage of me" is substantially lower in the UK (5%) and the US (6%) than in Morocco (20%) and Turkey (13%). Consistent with this, generalized trust also appears higher in WEIRD compared to non-WEIRD societies (see Fig. 1c in Section 3). Thus, together with behavioural evidence that trust – as a belief in others' benevolent motives – has been shown to correlate positively with cooperative behaviour in public goods games (Gächter et al. (2004); Balliet and Van Lange (2013); Kocher et al. (2015)) we expect that beliefs about others' cooperation are lower in Morocco and Turkey than in the UK and the US.

[4] Of the 18 studies considered in the analysis of Thöni and Volk (2018), 16 were conducted in WEIRD societies (USA, UK, Switzerland, Austria, The Netherlands, Denmark); 1 study included Japan (Kocher et al. (2008)); 1 study was done in Russia (Herrmann and Thöni (2009)). Studies not included in Thöni and Volk (2018) are, e.g., Kocher et al. (2015) (Germany); Bigoni et al. (2019) (Italy); Rustagi et al. (2010) (Ethiopia); and Martinsson et al. (2013) (Colombia and Vietnam).

[5] Bluffstone et al. (2020) qualitatively replicate these results using a standard public goods game without eliciting preferences for conditional cooperation.

With regard to our methodology to study our first behavioural mechanism, a related paper is Bigoni et al. (2019) who study a North-South divide in cooperation in Italy (Bigoni et al. (2016)). They found that preferences for conditional cooperation were similar between participants from the North and the South. Beliefs about others' cooperation, however, were higher in the North than in the South. Despite its North-South divide, Italy is one country, with internal migration and a shared linguistic, cultural, and institutional heritage. Our societies arguably differ more between each other than Italy differs internally.[6] Therefore, our study extends the evidence about the scope of cross-societal differences in the behavioural mechanisms underpinning voluntary cooperation by focusing on two countries that are culturally among the most distant from WEIRD countries. Furthermore, it also focuses on the role of punishment.

In summary, evidence on conditionally cooperative dispositions predominantly comes from WEIRD societies but results from the few non-WEIRD studies are comparable.[7] However, none of these studies conducted a systematic comparison of cooperative dispositions in WEIRD societies and societies that are culturally distant from them within one comparable design, nor do they elicit beliefs about others' cooperativeness and the willingness to punish other group members. Thus, together with the ambiguous predictions from the existing evidence on betrayal aversion and preferences for positive reciprocity, we expect little differences in conditionally cooperative dispositions between our four subject pools (Hypothesis 1a), but higher cooperative beliefs in the two WEIRD countries (Hypothesis 1b) which will lead to higher cooperation in WEIRD countries (Hypothesis 1c).

The second behavioural mechanism concerns *punishment*. Closest to our research question on this dimension are Herrmann et al. (2008) and Gächter and Herrmann (2009) who also studied cooperation and punishment, but neither elicited cooperative dispositions, nor beliefs. Herrmann et al. (2008) used a repeated game setting in fifteen (WEIRD and non-WEIRD) countries around the world. They found little differences in how people punish free riders, but strong differences in "antisocial punishment", that is, how people punish those who contribute the same or more than them. Antisocial punishment was largely absent in their UK and US subject pools but was substantial in their Turkish and Arabic subject pools (see also

[6] For other within-country (culture) experiments measuring social preferences, see, e.g., Brosig-Koch et al. (2011), Kim et al. (2017) and Choi et al. (2020).

[7] Comparing the rate of cooperative dispositions across the 20 WEIRD and 5 non-WEIRD samples included in Rustagi et al. (2010), Martinsson et al. (2013) and Thöni and Volk (2018) shows a similar percentage of conditional cooperators ($M_{\text{WEIRD}} = 59\%$; $M_{\text{non-WEIRD}} = 51\%$; Mann-Whitney $Z = 1.50$, $p = 0.148$) and a significantly higher percentage of free riders in WEIRD societies ($M_{\text{WEIRD}} = 12\%$; $M_{\text{non-WEIRD}} = 21\%$; Mann-Whitney $Z = 1.97$, $p = 0.048$).

Bruhin et al. (2020) for a further analysis of this result). Likely explanations include exposure to persistent resource scarcity (Prediger et al. (2014)) or strategic revenge (e.g., Sylwester et al. (2013)), which was possible given their repeated game design. Our one-shot design excludes this possibility.

If antisocial punishment is not only a (culture-specific) act of revenge in a repeated interaction, but also a "preference", it might be observed even in our one-shot design. The one-shot experiments of Gächter and Herrmann (2011) in Russia, and Gächter and Herrmann (2009) in Russia and Switzerland, suggest this possibility: Antisocial punishment did occur in their Russian subject pools, but was negligible in their Swiss groups. If these results generalize, then we might observe no differences between subject pools regarding punishment of free riders, and more antisocial punishment in Morocco and Turkey than in the US and the UK.

With regard to cooperation in the presence of punishment, the results from Gächter and Herrmann (2009) showed that cooperation was higher in Switzerland than in Russia. Similarly, Herrmann et al. (2008) found that cooperation was higher in subject pools with lower antisocial punishment. Together with Hypotheses 1a and 1b we therefore expect that cooperation in the presence of punishment will be higher in the WEIRD than in the non-WEIRD subject pools.

To summarize our hypotheses: We expect no subject-pool differences in the expected and actual punishment of free riders, but we expect lower antisocial punishment in the WEIRD compared to the non-WEIRD subject pools (Hypothesis 2a). In the presence of punishment, we expect beliefs and contributions to be higher in the WEIRD compared to the non-WEIRD subject pools (Hypothesis 2b).

3. Conceptual background and methods

3.1. Conceptual background of the behavioural mechanisms of voluntary cooperation

Our conceptual framework builds on Fischbacher and Gächter (2010); Gächter et al. (2017); and Weber et al. (2018). It models cooperation among a group of strangers in a public goods game: Individual i who lives in society s decides his or her contribution c_i^s as a function of his or her cooperative disposition a_i^s and beliefs b_i^s about others' average contribution: $c_i^s = f(a_i^s(b_i^s))$. This implies that voluntary cooperation levels between societies might differ because a_i^s or b_i^s differ between societies. Similarly, i's preferences for punishment of another individual j, $p_{ij}^s(c_i^s, c_j^s)$, depends on i's and j's contributions and might be society (culture) specific (Herrmann et al. (2008); Gächter et al. (2010)). Based on the results of Weber et al.

(2018), we assume that one's preference for punishment (p_{ij}^s) is independent of one's cooperative disposition (a_i^s).

Long-run historical processes, e.g., whether society s gravitates towards an "individualistic" or "collectivistic" culture (e.g., Greif (1994)), can shape the psychological dispositions a_i^s and p_{ij}^s and beliefs b_i^s by which people approach cooperation. Historical processes also shape the institutions around which society *s* organizes its economic and social life (Schulz et al. (2019); Schulz (2020); Henrich (2020); Rustagi (2020)). Beliefs b_i^s take a central role because actual experiences with institutions, combined with processes of cultural and parental transmission of values (e.g., Bisin and Verdier (2011); Hauk and Saez-Marti (2002); Guiso et al. (2008); Tabellini (2008); Henrich (2015)), will likely shape beliefs b_i^s. While we do not investigate the causal role of historical, cultural or institutional factors on beliefs and dispositions, by choosing societies that vary substantially along cultural dimensions we learn about the relative importance of beliefs and disposition in accounting for differences in voluntary cooperation.

Our one-shot experiments are designed to provide measures of the key variables of interest while excluding any strategic incentives that might be present in repeated games. These are: cooperative attitudes (dispositions) a_i^s, beliefs b_i^s, contributions c_i^s, punishment p_{ij}^s, and the expected punishment of i by j, $e(p_{ji}^s)$, from four societies *s*: Two societies are culturally similar WEIRD societies (UK & US) and two are culturally distant from the UK and US, but also similar to each other (Morocco & Turkey).

All subjects participate in two experiments, the dispositions-elicitation experiment (D-Game) followed by the punishment experiment (P-Game). The D-Game measures the first behavioural mechanism $c_i^s = f(a_i^s(b_i^s))$ in the absence of punishment. The P-Game provides the data for our second behavioural mechanism, which measures $c_i^s = f(a_i^s(b_i^s), e(p_{ji}^s))$. In the next section we describe the rationale for our selection of countries and the evidence about their cultural distance. Design details are in Sections 3.4 and 3.5.

3.2. Institutional and cultural background of the four subject pools

To establish cultural similarity and distance, we proceed in two steps. First, we rely on established cross-cultural and institutional indicators. Second, we use a new dataset to establish a quantitative summary statistic of cultural distance between our subject pools.

Figure 1 illustrates our first step. It shows that our subject pools come from similar societies within their respective WEIRD and non-WEIRD cluster regarding twelve frequently used quantitative characterizations of societies. On eleven of the twelve dimensions, the two WEIRD subject pools are culturally distant from the two non-WEIRD pools. Panels (*a*) and (*b*) show four indicators often used in cross-cultural research (value orientations and cultural dimensions); panels (*c*) and (*d*) display four indicators of the quality of institutions (government effectiveness; prevalence of rule violations; rule of law; and GDP per capita); and panels (*e*) and (*f*) depict trust and norms of civic cooperation, and preferences for negative and positive reciprocity, with the latter variable being the only one that is not culturally clustered. Further details and data sources are in the notes to Figure 1, and the Online Appendix Table A.1.

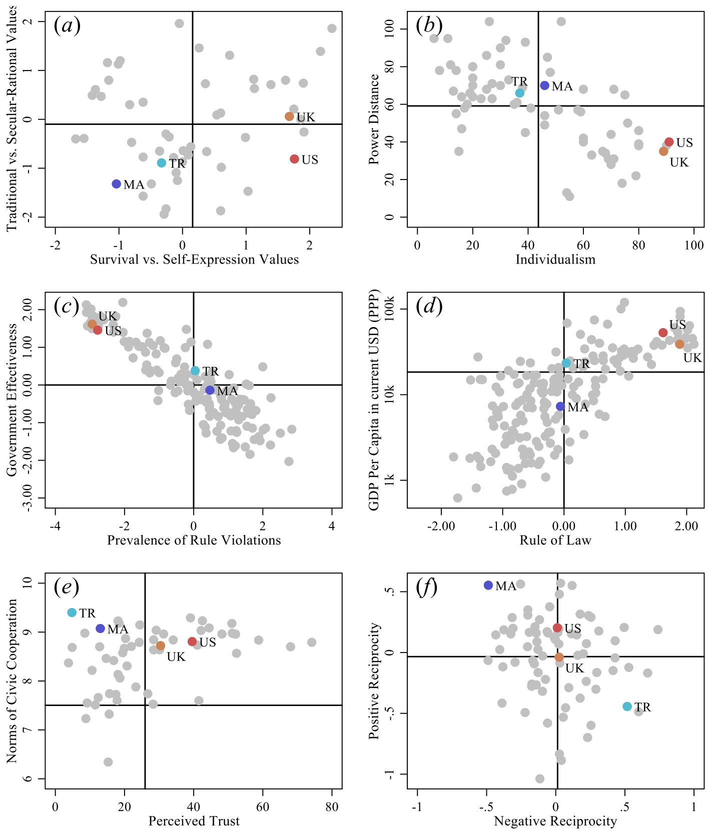


FIGURE 1. Institutional and cultural indicators for the United States (US), United Kingdom (UK), Morocco (MA) and Turkey (TR). The grey dots represent all other countries for which data was available. The black lines indicate the sample average values. Panel (*a*): *Traditional vs. Secular-Rational Values* (−1.94 = strongest emphasis on traditional values; 1.96 = strongest emphasis on secular-rational values) and *Survival vs. Self-Expression Values* (−1.68 = strongest emphasis on survival values; 2.35 = strongest emphasis on self-expression values) from Inglehart and Welzel (2005); Panel (*b*): Cultural dimensions *Power Distance* (11 = lowest; 104 = highest) and *Individualism* (6 = least individualist; 91 = most individualist) from Hofstede and Hofstede (2001); Panel (*c*): *Government Effectiveness* (−2.48 = lowest government effectiveness; 2.19 = highest government effectiveness) from the 2014 Worldwide Governance Indicators (Kaufmann and Kraay (2016)) and *Prevalence of Rule Violations* (−3.10 = lowest; 2.84 = highest) from Gächter and Schulz (2016); Panel (*d*): World Bank *GDP Per Capita* in current USD (PPP) for 2015 and *Rule of Law* (−2.39 = lowest rule of law; 2.12 = highest rule of law) from the 2014 Worldwide Governance Indicators (Kaufmann and Kraay (2016)); Panel (*e*): *Norms of Civic Cooperation* (1 = very weak norm of civic cooperation; 10 = very strong norm of civic cooperation) (Knack and Keefer (1997)), and *Perceived Trust*: percent of people who agree with the statement "most people can be trusted" by Inglehart et al. (2014); Panel (*f*): *Positive Reciprocity* (−1.04 = lowest positive reciprocity; 0.57 = highest positive reciprocity) and *Negative Reciprocity* (−0.49 = lowest negative reciprocity; 0.74 = highest negative reciprocity) from the Global Preference Survey (Falk et al. (2018)).

Our second step is a quantitative measure to characterize the *cultural distance* between the countries of our subject pools. This acknowledges the fact that WEIRD and non-WEIRD is a continuum and not a dichotomy (Henrich et al. (2010); Apicella et al. (2020)). To calculate cultural distance, we turn to Muthukrishna et al. (2020) who provide the most recent and most comprehensive data available. Using data from about 170,000 individuals from 80 countries, which represent approximately 85% of the world population, Muthukrishna *et al*. developed a quantitative measure of cultural distance between these 80 societies. Cultural distance is calculated as the ratio of the between-country and within-country variance in bilateral country comparisons of individual responses to questions in the World Values Survey (Waves 2005-09 and 2010-14). Questions are about a wide range of values and beliefs (e.g., attitudes to private affairs; family and religious values; attitudes to work; and political values).[8]

The quantitative measures of cultural distance between any pair of countries provide further evidence for the two distinct clusters of the societies we selected. In terms of ranked cultural distance using the US as the baseline, the UK is the 6th closest country to the US; Turkey is ranked 47th; and Morocco is ranked 59th (Egypt is ranked 80th, that is, farthest away from the US). Taking Morocco as the baseline shows that Turkey is ranked 18th; the US is ranked 60th, and the UK is ranked 71st (farthest away is Sweden, ranked 80th).[9]

From this diverse set of indicators, we conclude that our WEIRD and non-WEIRD subject pools come from two distant clusters, but share cultural similarity within their respective cluster. Thus, to the extent that we see differences in conditional cooperation, beliefs, and punishment between our WEIRD and non-WEIRD subject pools, we expect them to be bigger between the clusters than within them.

---

[8] Muthukrishna et al. (2020) calculated cultural distance analogous to measures of the genetic distance of populations (e.g., Cavalli-Sforza et al. (1994)). Cultural distance is calculated it as $(H_T - H_g)/H_T$ where H_T is the pairwise differences in answers of the WVS questions between two countries and H_g is the average number of pairwise differences in answers between two individuals from the same country. A cultural distance of 0 therefore means that countries share the exact same cultural values and beliefs; a value of 1 would imply two homogenous countries with entirely separate cultural values and beliefs.

[9] The data are taken from http://culturaldistance.com. Taking the US as the baseline, the cultural distance to the UK is 0.05; whereas the cultural distance of the US to Turkey it is 0.12 and to Morocco it is 0.15. The country with the largest cultural distance from the US is Egypt with a distance of 0.23. Taking Morocco as the baseline shows that the cultural distance of Morocco to Turkey is 0.06; the cultural distance to the US is 0.15 and to the UK it is 0.23. The country with the largest cultural distance from Morocco is Sweden with a distance of 0.34. See Figure A.1 in the Online Appendix for further details. To put these numbers into perspective, note that the largest cultural distance in Muthukrishna *et al*.'s data set is 0.57 (between Egypt and Norway), and the smallest is 0.01 (between Ukraine and Russia).

3.3. Participants and procedures

We follow Herrmann et al. (2008) in choosing student samples with comparable socio-economic status within their respective society. This methodology ensures subject pool comparability by minimising confounds to our variables of interest that come from the differing socio-demographic composition of the respective subject pools (see also Gächter (2010); Thöni (2019) for conceptual discussions of this point). We measure additional variation in our post-experimental questionnaire and use them as controls in our regression analyses.

A total of 388 students participated in our study; 93.8% of participants were nationals of their respective country.[10] We recruited 128 students at Stony Brook University in the US; 92 students at the University of Nottingham in the UK; 80 students at the École Nationale d'Agriculture de Meknès in Morocco; and 88 students at Istanbul Bilgi University in Turkey. Table 1 summarizes the characteristics of our samples.

TABLE 1. Characteristics of the four samples.

	US	UK	Morocco	Turkey
Average age (*SD*)	19.56 (2.75)	19.63 (1.85)	20.84 (1.44)	22.02 (1.93)
Female	59%	56%	65%	40%
Business and economics	9%	23%	1%	22%
Urban background	46%	63%	48%	67%
Average number of siblings (*SD*)	1.69 (1.35)	1.69 (1.37)	3.03 (2.18)	1.79 (2.01)
Middle class	58%	84%	89%	42%
No. obs.	106	88	80	86
Average payoff in local currency (*SD*)	USD 13.87 (3.32)	GBP 11.50 (1.53)	MAD 100.46 (21.64)	TRY 40.93 (7.95)

Notes: Business and Economics: percentage of participants studying business or economics. Urban background: percentage of participants who lived most of their life in a town with at least 10,000 inhabitants. Middle class: percentage of participants who self-reported their family income at age sixteen to be at least average compared to other families.

The experiments were computerized and conducted with z-Tree (Fischbacher (2007)). At the École Nationale d'Agriculture de Meknès the experimenters recruited participants from the local student population. At the other three universities, we used ORSEE (Greiner (2015)) to recruit participants. In all samples, subjects could only participate once in the experiment.

[10] We excluded 28 participants from the sample of 388 students (22 participants from the US, 4 participants from the UK and 2 participants from Turkey) who indicated in the post-experimental questionnaire that they were not citizens of the respective countries. All main results of this paper hold independently of their inclusion.

We follow the established rules for conducting cross-cultural economic experiments (Roth et al. (1991); Thöni (2019)). In all four laboratories, the experiments were run by local research assistants. The instructions and software were presented in the local language.[11] The instructions used neutral language. Participants made their decisions in private with visual separations between workstations. The experimental sessions lasted approximately 90 minutes and were conducted according to a strict protocol to minimize the differences in the way sessions were run across countries. We paid participants in private at the end of each session. The stake sizes reflect local purchasing power and student wages so that, in real terms, the possible real earnings were similar across subject pools.

Each session consisted of two experimental games (the D-Game followed by the P-Game) and a socio-economic questionnaire. Participants were randomly re-matched after the D-Game. We did not provide any feedback after the D-Game to prevent participants from updating their beliefs and to reduce potential income effects and strategic play in the P-Game. However, as will become clear in Section 3.5, in contrast to the D-Game, the P-Game requires feedback on others' contribution, which is why the P-Game always followed the D-Game.

3.4. Disposition elicitation game (D-Game)

To measure cooperative dispositions, we used the strategy method developed by Fischbacher et al. (2001). Participants were randomly allocated to groups of four. Each received an endowment of 20 tokens and decided how much to contribute to a common "project". Individual payoffs were determined by the following function:

$$\pi_i = 20 - g_i + 0.4 \cdot \sum_{j=1}^{4} g_j. \qquad (1)$$

Participant i's contribution to the public good is given by g_i, and the size of the public good is the sum of all contributions. The marginal per capita return of the public good is 0.4 monetary units (MU). Although the social optimum entails full contributions, the individually money-maximizing strategy is to contribute nothing.

Participants made two decisions in the D-Game: they chose an *unconditional contribution* to the project and filled in a *contribution table*. This table allowed for conditioning the participant's own contribution on the rounded average contribution of her three other group members. To ensure incentive compatibility, the actual contribution for one randomly chosen participant per group was taken from their contribution table according to the average

[11] The instructions are available in Online Appendices J-K. We translated the instructions from English into Arabic or Turkish and had them back translated by another person in order to make sure that instructions were as identical as possible (Brislin (1970)).

contributions of the three co-players. After making contribution decisions, we elicited participants' beliefs about the other group members' average unconditional contribution. Participants earned three MUs for guessing correctly, two MUs for a deviation of one point, one MU for a deviation of two points and zero MUs for a higher deviation (Gächter and Renner (2010)).[12] We did not provide feedback after the D-Game and fixed this sequence across all sessions to prevent participants' from updating their beliefs about other participants' behaviour.

Using the criteria outlined in Thöni and Volk (2018), we classify participants into different types of cooperative dispositions according to their contribution tables. *Conditional Cooperators* (CC) show a positive correlation between their own conditional contributions and the average contributions of their fellow group members (i.e., Pearson's $\rho \geq 0.5$) or at least one increase in their contribution schedule. *Free Riders* (FR) contribute nothing for every possible average contribution of their group members. We refer to participants who are not classified as either CC or FR as *Unclassified Others* (OT). The results from the D-Game provide the data for our first behavioural mechanism $c_i^s = f(a_i^s(b_i^s))$ presented in Section 4.

3.5. Punishment game (P-Game)

The P-Game consists of two stages, both of which were fully explained in the pre-play instructions: First, participants choose their contribution in a one-shot public goods game for which the individual payoff function is given by (1). Then participants indicate their beliefs about the other group members' average contribution to the public good. We did not incentivize beliefs in the P-Game in order to avoid punishment motivated by disappointment due to wrong beliefs or income effects (Cubitt et al. (2011)). In the second stage of the game, participants learned their group members' individual contributions and could assign up to five punishment points to each. Each assigned punishment point cost the punisher one MU and removed two MUs of the targeted person's income. Individual payoffs were determined by the following function:

$$\pi_i = \pi_i^{S1} - \sum_{j=1}^{4} p_{ij} - 2 \cdot \sum_{j=1}^{4} p_{ji}. \qquad (2)$$

Participant i's payoff from the first stage is given by π_i^{S1}. The punishment points group member i allocates to group member j are denoted by p_{ij}, and p_{ji} denotes the punishment points allocated by j to i. Each participant received 10 MUs to cover potential losses. In the P-Game, the punishment $e(p_{ji}^s)$ that group member i expects to receive might influence i's contribution.

[12] Belief hedging might be a concern in this design. However, we believe this to be unlikely for two reasons. First, incentives to hedge are minor here. Second, Blanco et al. (2010) find no evidence for belief hedging in a similar social dilemma experiment.

After allocating punishment points to their group members, participants were asked to state how many punishment points they expected to receive from each of their fellow group members. The results from the P-Game provide the data for our second behavioural mechanism $c_i^s = f(a_i^s(b_i^s), e(p_{ji}^s))$ presented in Section 5.

4. Behavioural Mechanism I: Preferences for conditional cooperation and beliefs

4.1. Preferences for conditional cooperation (“dispositions”)

After classifying the participants according to their individual cooperative dispositions (see Section 3.4), we compare the distributions of cooperative dispositions across societies in Figure 2a (a_i^s in our conceptual framework).[13] Using χ^2-tests, we find that the distributions of cooperative dispositions are statistically significantly different across the four samples ($p = 0.043$). The most common cooperative disposition in all four samples is conditional cooperation, with a statistically similar share of CC across the four subject pools ($\chi^2(3) = 1.70$, $p = 0.637$). However, the shares of FR ($\chi^2(3) = 9.19$, $p = 0.027$) and the shares of OT ($\chi^2(3) = 6.67$, $p = 0.083$) vary across the four samples.[14]

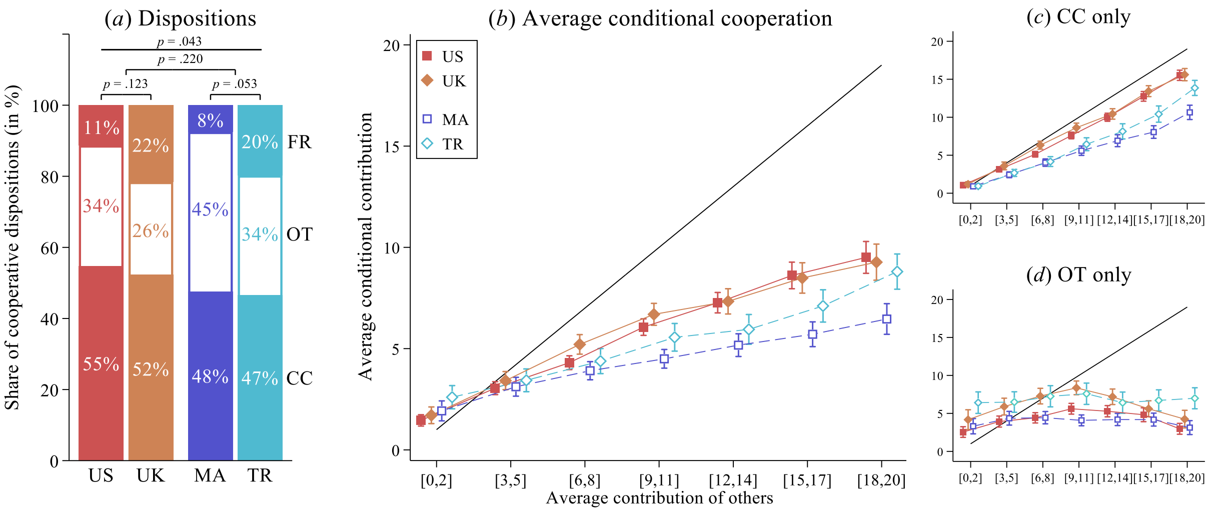


FIGURE 2. Preferences for conditional cooperation (a_i^s in our conceptual framework). Panel (*a*): The share of conditional cooperators (CC), free riders (FR) and others (OT); *p*-values from a (pooled) χ^2 tests. Panel (*b*): Average contribution schedule by country in the D-Game; all participants. Panel (*c*): Conditional cooperators (CC) only. Panel (*d*): Others (OT) only. The 45°-line corresponds to perfect conditional cooperation. The error bars indicate ±1 *SEM*.

---

[13] The results reported here are robust to an alternative classification (see Online Appendix B).

[14] CC include perfect conditional cooperators, who equal the others’ contribution (13 in the US, 9 in the UK, 3 in MA, and 3 in TR). OT include unconditional cooperators (3 in the US, 2 in the UK, 0 in MA, 9 in TR), triangle contributors (24 in the US, 16 in the UK, 23 in MA, 6 in TR) and unclassified others (9 in the US, 5 in the UK, 13 in MA, 14 in TR).

Comparing the WEIRD cluster (pooling the US and UK samples) to the non-WEIRD cluster (pooling the Morocco and Turkey samples) reveals statistically similar distributions of cooperative dispositions ($p = 0.220$). Within clusters, there are no significant differences in the distribution of cooperative dispositions between the US and UK samples ($p = 0.123$) and weakly significant differences when comparing the distributions of the Morocco (MA) and Turkey (TR) samples ($p = 0.053$).

Another way of investigating differences in cooperative dispositions across samples is to compare the *extent* of conditional cooperativeness. We use the average contributions in the contribution table of the D-Game as a measure of conditional cooperativeness (Figure 2b). Across the four samples, on average participants increase their conditional contributions as the average contribution of others rises, but they only imperfectly match the average contribution of others by undercutting others' contribution. Imperfect conditional cooperation results in average conditional contributions below the 45°-line (which corresponds to perfect conditional cooperation). Figure 2b also reveals that societal differences in average conditional contributions are small in the "behaviourally relevant" interval between 0 and 11 where 91% of the average unconditional contributions of other group members fall. Figure 2c illustrates average conditional cooperation for conditional cooperators (CC) only, and Figure 2d for others (OT) only.

Finally, we use a regression analysis to test for across and within-cluster differences in the level and slope of the participants' contribution schedules (Table 2).[15] We refer to the UK and the US as the WEIRD cluster and to Morocco and Turkey as the non-WEIRD cluster. First, we investigate across-cluster differences in a pooled estimation by regressing the conditional contribution on the average contribution of other group members, a dummy variable for the non-WEIRD cluster and an interaction term between the dummy and the average contribution of others to measure differences in slopes (Table 2, Col. 1). Additionally, we control for the socio-economic background of participants.

[15] When estimating the same regression models including only CC, we find slight variation in conditional contributions across the two clusters and within the non-WEIRD cluster (Table C.1. in the Online Appendix). Using the same method to compare the conditional contributions of OT shows only significant level differences within the WEIRD cluster and overall a very poor fit of the regression model (Table C.2. in the Online Appendix).

TABLE 2. Conditional contributions across subject pools.

Dependent variable: conditional contribution	(1)	(2)
Average contributions of others	0.438*** (0.034)	0.456*** (0.044)
Average contribution of others × Non-WEIRD cluster	−0.152*** (0.045)	
Non-WEIRD cluster	0.653 (0.509)	
Average contribution of others × UK		−0.040 (0.068)
UK		0.851 (0.588)
Average contribution of others × MA		−0.216*** (0.060)
MA		1.021* (0.593)
Average contribution of others × TR		−0.127** (0.063)
TR		1.171 (0.755)
Socio-economic controls	Yes	Yes
Constant	3.016 (2.655)	3.062 (2.710)
R^2	0.14	0.15
No. obs. (Clusters)	7560 (360)	7560 (360)

Notes: OLS estimation with robust *SE* clustered on individuals in parentheses. The dependent variable is the conditional contribution elicited on the contribution table in the D-Game where participants indicated their own conditional contribution for each of the 21 average contributions of their group members. WEIRD cluster: UK and US; non-WEIRD cluster: Morocco and Turkey. Socio-economic control variables: age, female, urban background, middle class, single child, economics/business student. None is significant at any conventional level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

The coefficient for the average contributions of others is positive and highly significant, but smaller than 1 implying imperfect conditional cooperation in the WEIRD cluster. Conditional contributions increase on average by 0.438 for each additional money unit contributed. The negative and highly significant interaction term implies a smaller increase in conditional contributions in the non-WEIRD cluster. The statistically insignificant non-WEIRD dummy indicates a similar average conditional contribution in both clusters.[16]

[16] The estimation results for the UK (Model (2) in Table 3) are similar to those from Weber et al. (2018) who also ran a comparable strategy method public goods game with UK subjects. In their data, the coefficient (*SE*) for average contribution of others is 0.401 (0.022).

Next, we estimate a similar model with country dummies and country-specific interaction terms to test for within-cluster variation (Table 2, Col. 2). Here, the US sample serves as the comparison group. The statistical insignificance of the UK dummy and its interaction term "Average contribution of others × UK" imply no variation in the slope or levels within the WEIRD cluster comparing the US and UK samples. Additionally, we find no statistically significant differences in the coefficients of "Average contribution of others × MA" and "Average contribution of others × TR" ($F(1, 359) = 2.16$, $p = 0.142$) or the MA and TR dummies ($F(1, 359) = 0.03$, $p = 0.858$). This indicates similar slopes and levels of conditional contributions within the non-WEIRD cluster.

We summarize our findings on preferences for conditional cooperation as **Result 1a:** *We find a similar distribution of cooperative dispositions across our four subject pools, but a slightly higher conditional cooperativeness in the WEIRD compared to the non-WEIRD cluster.*

Result 1a provides mixed support for Hypothesis 1a, which stipulates no differences between subject pools. While the distribution of cooperative dispositions is similar across subject pools, we do find slightly weaker conditional cooperativeness in Morocco and Turkey than in the UK and US. Our next step is to test Hypothesis 1b that beliefs about others' cooperation are higher in the UK and US than in Morocco and Turkey.

4.2. Beliefs and contributions

Figure 3a shows *beliefs* about unconditional contributions in the D-Game (b_i^s in our conceptual framework). We find significant variation in beliefs across the four samples (Kruskal-Wallis test, $\chi^2(3) = 18.58$, $p < 0.001$). Beliefs are significantly higher in the WEIRD cluster compared to the non-WEIRD cluster. We do not find significant differences in beliefs within the WEIRD cluster, that is, comparing the US and UK samples. For the non-WEIRD cluster, we find highly significant differences between the MA and TR samples. The between-cluster difference in beliefs is driven by lower beliefs in the MA sample with beliefs in the TR sample not being significantly different from those held in the US or UK sample (Mann-Whitney $Z = -0.52$, $p = 0.602$; $Z = -1.11$, $p = 0.268$; resp.).[17]

We collect these findings in **Result 1b**, which confirms our Hypothesis 1b: *Beliefs about others' cooperation are higher in the WEIRD compared to the non-WEIRD cluster.*

[17] We also test for differences in the accuracy of beliefs—defined as deviation from the actual contribution of others—and find a statistically similar belief accuracy across samples (overall average 1.8, Kruskal-Wallis test, $\chi^2(3) = 5.57$, $p = 0.134$; for details see Table D.1 in the Online Appendix).

We now turn to cooperation and Hypothesis 1c which predicts that cooperation is higher in the WEIRD cluster than in the non-WEIRD cluster. Figure 3b illustrates the average unconditional *contributions* in the D-Game (c_i^s in our conceptual framework) which vary significantly across the four samples (Kruskal-Wallis test, $\chi^2(3) = 19.62$, $p < 0.001$). Unconditional contributions in the WEIRD cluster are significantly higher than in the non-WEIRD cluster. There are no differences in unconditional contributions within-clusters when comparing the US and UK samples or the MA and TR samples.

Result 1c summarizes this finding: *Unconditional contributions are higher in the WEIRD than in the non-WEIRD cluster.*

This supports Hypothesis 1c. Our next step is to explain contributions as a function of dispositions and beliefs.

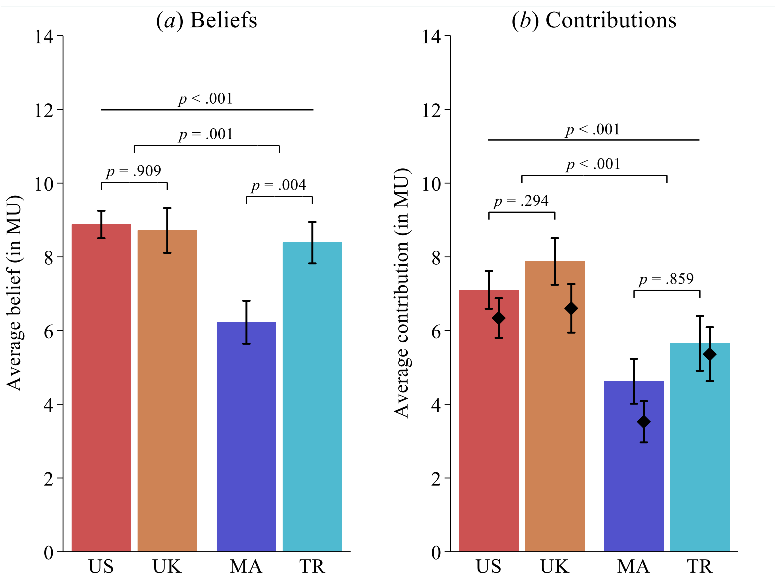


FIGURE 3. Average beliefs (Panel *a*); and average unconditional contributions to the public good (Panel *b*) in the D-Game by country. Diamonds indicate the average predicted contributions based on dispositions and beliefs (described in section 4.3). The error bars indicate ±1 *SEM*; *p*-values from Kruskal-Wallis and (pooled) Mann-Whitney tests.

4.3. Explaining contributions as a function of dispositions and beliefs

To explain contributions, we calculate "Predicted contribution" by combining a participant's individual contribution table and their belief about others' contributions ("Belief") as elicited in the D-Game. Formally put, we calculate for each individual *i* their predicted contribution $\widehat{c_i}$ as $a_i(b_i) \rightarrow \widehat{c_i}$ where a_i is *i*'s contribution schedule as a function of others' average contributions, and b_i is *i*'s belief about others' average contribution. The sample average predicted contributions $\widehat{c^s}$ are depicted as diamonds in Figure 3b.

Predicted contributions vary significantly across the four samples (Kruskal-Wallis test, $\chi^2(3) = 15.03$, $p = 0.002$) and tend to be lower than actual contributions (Figure 3b). Predicted contributions in the WEIRD cluster are significantly higher than in the non-WEIRD cluster (pooled Mann-Whitney $Z = 3.73$, $p < 0.001$). There are no differences in our within-cluster analysis of predicted contributions when comparing the US and UK samples (Mann-Whitney $Z = 0.09$, $p = 0.931$) or the MA and TR samples (Mann-Whitney $Z = -0.65$, $p = 0.517$). Yet, the predicted contributions are significantly lower compared to the actual contributions for all but the Turkey samples (Wilcoxon signed-rank test $Z_{US} = 2.06$, $p_{US} = 0.040$; $Z_{UK} = 2.88$, $p_{UK} = 0.004$; $Z_{MA} = 1.52$, $p_{MA} = 0.131$; $Z_{TR} = 1.89$, $p_{TR} = 0.059$).[18]

To explain the gap between actual and predicted contributions, we follow Fischbacher and Gächter (2010) and regress unconditional contributions (c_i^s) on predicted contributions ($\widehat{c_i^s} = a_i^s(b_i^s)$), beliefs ($b_i^s$), a non-WEIRD cluster dummy and interaction terms in a pooled model (Table 3, Col. 1). We also control for the participants' socio-economic background.

The coefficients "Predicted contribution" and "Belief" are both positive and highly significant. The results show no significant level differences in contributions across clusters (non-WEIRD cluster $b = 0.164$, $p = 0.850$) when controlling for beliefs, suggesting that the observed differences in behaviour might be driven by variation in beliefs. There are also no differences in the way "Predicted contributions" and "Belief" are associated with the contribution behaviour across the clusters. This follows from the statistically insignificant dummy variable "Non-WEIRD cluster", as well as interaction terms "Predicted contribution × Non-WEIRD cluster" and "Belief × Non-WEIRD cluster".

The significant influence of beliefs on contributions across both clusters shows that conditional cooperation is greater than the level implied by "Predicted contribution" alone, that is, "Belief" matters on top of "Predicted contribution", a finding that Fischbacher and Gächter (2010) also report.[19]

[18] We also test for the accuracy of "Predicted contribution" across samples. Comparing predicted with actual contributions in the D-Game yields accuracy of more than 61% for all four societies and no significant differences in the accuracy of predicted contributions across subject pools (Figure E.1 in the Online Appendix).

[19] Reassuringly, the estimation results for the UK (Col. 1 in Table 3) are similar to those from Weber et al. (2018) who also ran a comparable strategy method public goods game with UK subjects. The coefficient (*SE*) for predicted contribution is 0.532 (0.067) and for beliefs it is 0.392 (0.068).

TABLE 3. Explaining unconditional contributions in the D-Game.

Dependent variable: conditional contribution	(1)	(2)
Predicted contribution	0.504*** (0.090)	0.463*** (0.112)
Belief	0.300*** (0.116)	0.397*** (0.141)
Predicted contribution × Non-WEIRD cluster	0.088 (0.150)	
Belief × Non-WEIRD cluster	−0.191 (0.159)	
Non-WEIRD cluster	0.164 (0.863)	
Predicted contribution × UK		0.049 (0.181)
Belief × UK		−0.127 (0.221)
UK		1.616 (1.061)
Predicted contribution × MA		0.225 (0.170)
Belief × MA		−0.536*** (0.174)
MA		2.652** (1.201)
Predicted contribution × TR		0.081 (0.190)
Belief × TR		−0.059 (0.210)
TR		−0.903 (1.144)
Socio-economic controls	Yes	Yes
Constant	2.042 (2.390)	1.026 (2.369)
R^2	0.49	0.50
No. obs.	360	360

Notes: OLS estimates with robust *SE* in parentheses. WEIRD cluster: UK and US; non-WEIRD cluster: Morocco and Turkey. Socio-economic control variables: age, female, urban background, middle class, single child, economics/business student. None is significant at any conventional level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Turning to the within-cluster comparisons (Table 3, Col. 2), within the WEIRD cluster, we find neither level differences in contributions between the US and UK samples, nor significant differences in the way “Predicted contribution” and “Belief” are associated with the contributions. To investigate differences in the non-WEIRD cluster, we test for significant differences in the coefficients of the MA and TR dummies, as well as the interaction terms with predicted contributions and beliefs. We find significant level differences in contributions ($F(1, 342) = 7.39, p = 0.007$) and a significant differences in the way beliefs are associated with contributions ($F(1, 342) = 6.79$, $p = 0.010$). Yet, we find no differences in the way "Predicted contributions" influence contribution behaviour across the MA and TR samples ($F(1, 342) = 0.51$, $p = 0.474$). This shows considerable within-cluster variation for the non-WEIRD cluster. Contribution levels are higher in the MA compared to the TR sample, but participants from the TR sample display a larger increase in their contributions for a higher expected average contribution of others.

The results of the first regression model have two important implications: First, when we control for beliefs, we find no significant level differences in contributions between the two clusters (the non-WEIRD cluster dummy is insignificant). Because "Belief" is highly significant, this suggests that differences in beliefs are an important driver of subject pool differences in cooperation. Second, the non-significant interactions between "Predicted contributions" and the cluster dummy indicate that a one-unit increase in "Predicted contributions" increases actual contributions by the same amount across clusters. The same interpretation holds for the non-significant interaction between "Belief" and the cluster dummy. This shows that the relative importance of the "Predicted contributions" and "Belief" in explaining behaviour is similar across clusters.

4.4. Discussion

In terms of our hypotheses about the first behavioural mechanism, we find mixed support for Hypothesis 1a that predicts similar distributions of conditionally cooperative dispositions a_i^s across the four subject pools. The distribution of conditionally cooperative dispositions previously mainly elicited in WEIRD subject pools (Thöni and Volk (2018)) generalizes to our non-WEIRD subject pools. However, average slopes are slightly lower in the non-WEIRD subject pools than in the WEIRD subject pools. The beliefs b_i^s about others' cooperation, are higher in the WEIRD cluster than in the non-WEIRD cluster, which confirms Hypothesis 1b. Beliefs appear as an important factor that explains contributions in both clusters including higher contributions in the WEIRD cluster than in the non-WEIRD cluster, thereby confirming Hypothesis 1c. In summary, in terms of our first behavioural mechanism ($c_i^s = f(a_i^s(b_i^s))$), cooperative dispositions and beliefs explain unconditional contributions in all subject pools.

5. Behavioural Mechanism II: Punishment

Our next step is to investigate the second behavioural mechanism—expected ($e(p_{ji}^{s})$) and actual punishment (p_{ij}^{s})—and its implication for cooperation and beliefs ($c_i^s = f(a_i^s(b_i^s), e(p_{ji}^s))$).[20] For this purpose, we analyse the P-Game described in Section 3.5 and test Hypotheses 2a and 2b.

5.1. Expected punishment

After participants made their contribution decisions in the P-Game and were informed about the contributions of their group members, we elicited expected punishment by asking participants to state the number of punishment points they expect to receive from each of their fellow group members. In applying the second behavioural mechanism described in Section 3.1, we analyse *expected* punishment separately for a negative or non-negative deviation from the potential punisher's contribution (Figure 4). Expected prosocial punishment (punishment of *negative* deviations from the punisher's contribution) differs significantly across the four subject pools (Kruskal-Wallis test, $\chi^2(3) = 58.42$, $p < 0.001$). We find highly significant differences in expected prosocial punishment across the WEIRD and non-WEIRD clusters. We also find substantial differences within the WEIRD cluster with significantly higher expected prosocial punishment in the UK than the US sample. Within the non-WEIRD cluster, we find a highly significant difference with higher expected prosocial punishment in the TR than the MA sample.

Expected antisocial punishment (punishment of *non-negative* deviations from the punisher's contribution) is statistically similar across the four samples (Kruskal-Wallis test, $\chi^2(3) = 5.56$, $p = 0.135$). It is also statistically similar when comparing the WEIRD and non-WEIRD clusters, as well as when comparing the US and UK samples or the MA and TR samples. The results of our non-parametric analysis are largely confirmed by Tobit regressions that control for the magnitude of the negative or positive contribution deviations and the subjects' socio-economic background (Online Appendix, Table G.1).[21]

[20] In Online Appendix F, we also test the assumption of our framework that p_{ij}^{s} is independent of a_i^s – see Section 3.1; Weber et al. (2018); Molleman et al. (2019) and find it confirmed.

[21] A further question is whether the accuracy of expected punishment differs across samples. We find a significantly lower accuracy of expected punishment in the WEIRD cluster compared to the non-WEIRD cluster, due to overestimation of the use of punishment in the US and UK samples. We also find a significant within-cluster differences in accuracy for the non-WEIRD cluster only (Figure H.1 in the Online Appendix).

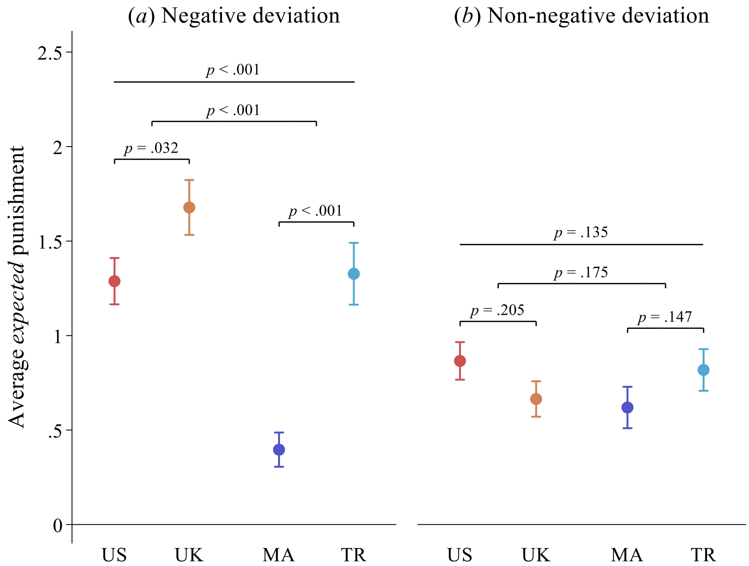


FIGURE 4. Average *expected* punishment points by country for (*a*) a negative deviation or (*b*) a non-negative deviation from the punisher's contribution. The error bars show ±1 *SEM* clustered on individuals; *p*-values from Kruskal-Wallis and (pooled) Mann-Whitney tests.

5.2. Actual punishment

Actual punishment indicates a preference ("a willingness to pay") for inflicting costs on a group member. Figure 5 shows the actual punishment distinguishing again between negative and non-negative deviations of the target from the punisher's contribution (i.e., prosocial and antisocial punishment). Actual prosocial punishment differs weakly significantly across the four subject pools (Kruskal-Wallis test, $\chi^2(3) = 6.92$, $p = 0.075$). Yet, we find statistically similar levels of actual prosocial punishment across the WEIRD and non-WEIRD clusters, and within the WEIRD cluster when comparing the US and UK samples. Within the non-WEIRD cluster, actual prosocial punishment is significantly higher in the TR than the MA sample.

Antisocial punishment appears statistically similar across the four samples (Kruskal-Wallis test, $\chi^2(3) = 1.40$, $p = 0.135$). Although the differences seem small, we find significantly higher antisocial punishment in the WEIRD than the non-WEIRD cluster. We neither find significant variation within the WEIRD cluster comparing the US and UK samples, nor within the non-WEIRD cluster comparing the MA and TR samples. A Tobit regressions analysis largely confirms these results (Online Appendix, Table G.2).

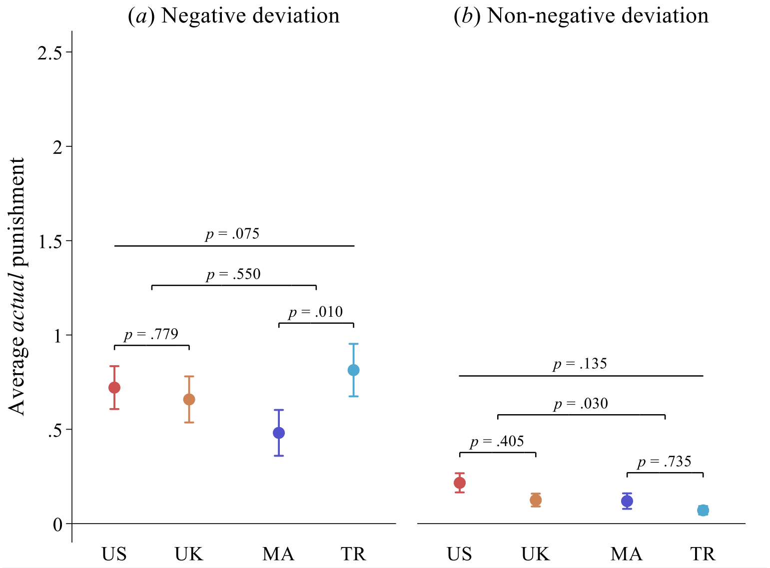


FIGURE 5. Average *actual* punishment points by country for (*a*) a negative deviation or (*b*) a non-negative deviation from the punisher's contribution. The error bars show ±1 *SEM* clustered on individuals; *p*-values from Kruskal-Wallis and (pooled) Mann-Whitney tests.

We summarize these findings in **Result 2a:** *Actual punishment of free riding is similar across the two clusters. Antisocial punishment is slightly higher in the WEIRD than the non-WEIRD subject pools.*

These results partially reject Hypothesis 2a. As predicted, we find no differences in actual punishment of free riding. However, contrary to the prediction of Hypothesis 2a, we find slightly lower antisocial punishment in the non-WEIRD subject pools than in the WEIRD subject pools. We will return to this surprising result in Section 7.

Our final step is to explain contributions in the P-Game and to test Hypothesis 2b which predicts higher cooperation in the WEIRD cluster than the non-WEIRD cluster.

5.3. Explaining contributions in the P-Game

Figure 6a illustrates average beliefs about others' contributions (b^s) per subject pool s in the P-Game. Beliefs differ significantly across the four subject pools (Kruskal-Wallis test, $\chi^2(3) = 17.93$, $p < 0.001$). Beliefs are significantly higher in the WEIRD cluster than in the non-WEIRD cluster. Within clusters, beliefs are weakly significantly higher in the UK than in the US sample and statistically similar in the MA and TR samples.[22]

[22] Accuracy of beliefs—defined as the deviation from the actual contribution of others ($c_i^s - b_i^s$)—across the four subject pools is higher in the WEIRD cluster than in the non-WEIRD cluster (average accuracy is 0.65 in the WEIRD cluster vs. 2.13 in the non-WEIRD cluster). The difference in inaccuracy between clusters is significant, whereas differences within clusters are not (Table D.1 in the Online Appendix).

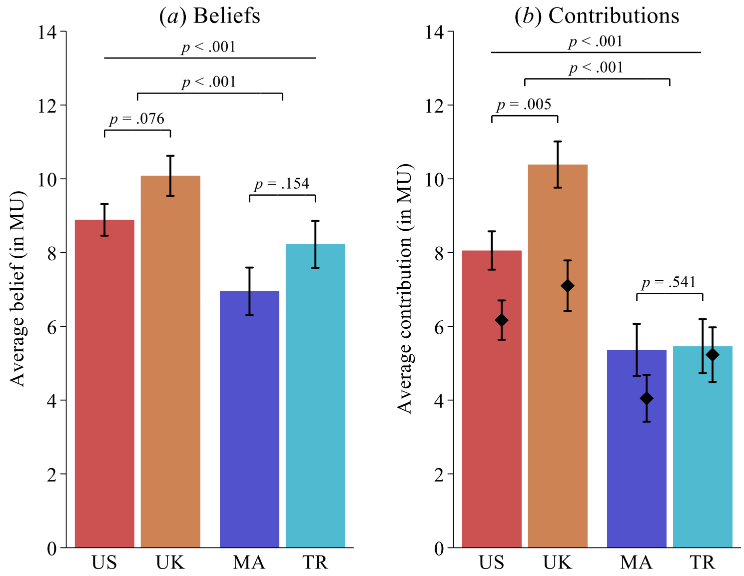


FIGURE 6. Average beliefs about others' contributions (*a*) and contributions to the public good (*b*) in the P-Game by country. Diamonds indicate the average predicted contributions based on dispositions (from the D-Game) and beliefs. The error bars indicate ±1 *SEM*; *p*-values from Kruskal-Wallis and (pooled) Mann-Whitney tests.

Figure 6b shows the average contributions in the P-Game which differ significantly across the four samples (Kruskal-Wallis test, $\chi^2(3) = 47.42$, $p < 0.001$). We also find significantly higher contributions in the WEIRD cluster compared to the non-WEIRD cluster. A comparison within the WEIRD cluster shows significantly higher contributions in the UK compared to the US sample, but we find no differences within the non-WEIRD cluster.

We also compare contributions in the D-Game with contributions in the P-Game. Interestingly, contributions in the P-Game are significantly higher than in the D-Game for the US (8.06 vs. 7.10, $Z_{US} = -2.42$, $p_{US} = 0.015$) and the UK (10.39 vs. 7.88, $Z_{UK} = -3.65$, $p_{UK} < 0.001$) but not for Morocco (5.36 vs. 4.63, $Z_{MA} = -0.78$, $p_{MA} = 0.438$) and Turkey (5.47 vs. 5.65, $Z_{TR} = -0.45$, $p_{TR} = 0.656$; all tests are Wilcoxon signed-rank tests).

We summarize these findings in **Result 2b**, which supports Hypothesis 2b: *In the presence of punishment, beliefs and contributions are higher in the WEIRD than in the non-WEIRD cluster.*

Next, we check if the effect of punishment on contributions goes via beliefs. To do that, we apply the first behavioural mechanism to explain contributions as a function of conditionally cooperative dispositions in the D-Game and beliefs in the P-Game ($c_i^s = f(a_i^s(b_i^s))$). The resulting predicted contributions (illustrated by diamonds in Figure 7b) vary significantly across the four samples (Kruskal-Wallis test, $\chi^2(3) = 13.83$, $p = 0.003$). Predicted contributions in the WEIRD cluster are significantly higher than in the non-WEIRD cluster (pooled Mann-Whitney $Z = 3.68$, $p < 0.001$). There are no differences in our within-cluster analysis of predicted contributions when comparing the UK and US samples (Mann-Whitney $Z = 0.91$, $p = 0.362$) or the MA and TR samples (Mann-Whitney $Z = -0.25$, $p = 0.806$). Yet, the predicted contributions are significantly lower than actual contributions for all samples (Wilcoxon signed-rank test $Z_{MA} = 1.73$, $p_{MA} = 0.084$; $Z_{TR} = 1.95$, $p_{TR} = 0.051$; $Z_{UK} = 4.89$, $p_{UK} < 0.001$; $Z_{US} = 4.02$, $p_{US} < 0.001$).

To explain the gap between actual and predicted contributions in the P-Game, we regress, in analogous fashion to Table 3, contributions (c_i^s) on predicted contributions ($\widehat{c_i^s} = a_i^s(b_i^s)$), beliefs ($b_i^s$), the non-WEIRD cluster dummy and interaction terms pooling the four samples (Table 4, Col. 1). The coefficients "Predicted contribution" and "Belief" are highly significant, and the coefficient for "Belief" appears larger than the coefficient in Table 3. This implies that, in the presence of punishment, beliefs are becoming more important for explaining contributions. Matching the expected contribution of others might be a strategy to avoid punishment in this game. The weakly significant positive coefficient "Predicted contribution × Non-WEIRD cluster" and the highly significant negative coefficient "Belief × Non-WEIRD cluster" indicate differences in the effect of predicted contributions and beliefs across clusters. The within-cluster analyses of the WEIRD cluster reveal no significant differences in the effect of predicted contributions, beliefs or level differences in conditional contributions when comparing the US and UK samples (insignificant interactions between "Predicted contributions" or "Belief" and the respective country dummy; insignificant country dummies; Table 4, Col. 2). Similarly, we find no significant within-cluster variation in the effect of predicted contributions, beliefs or level differences within the non-WEIRD cluster ($F(1, 342) = 0.00$, $p = 0.967$; $F(1, 342) = 0.03$, $p = 0.863$; $F(1, 342) = 1.03$, $p = 0.310$; resp.).

The analysis above does not capture the direct effect of punishment on contribution. A simple correlational analysis of expected punishment and beliefs or contributions might provide some insight on the effect of expected punishment on behaviour. In the WEIRD cluster, both expected contributions and actual contributions are positively associated with the average expected punishment (Spearman's $\rho = 0.37$, $p < 0.001$; $\rho = 0.19$, $p < 0.001$; resp.). In the non-WEIRD cluster, expected contribution and actual contribution are unrelated to the average expected punishment (Spearman's $\rho = 0.07$, $p = 0.369$; $\rho = -0.10$, $p = 0.194$; resp.). This suggests a stronger effect of punishment on behaviour in the WEIRD compared to the non-WEIRD cluster.

TABLE 4. Explaining unconditional contributions in the P-Game.

Dependent variable: conditional contribution	(1) By cluster	(2) By country
Predicted contribution	0.205*** (0.059)	0.277*** (0.089)
Belief	0.828*** (0.081)	0.732*** (0.125)
Predicted contribution × Non-WEIRD cluster	0.231* (0.120)	
Belief × Non-WEIRD cluster	−0.477*** (0.137)	
Non-WEIRD cluster	0.580 (0.858)	
Predicted contribution × UK		−0.147 (0.122)
Belief × UK		0.164 (0.176)
UK		0.540 (1.055)
Predicted contribution × MA		0.160 (0.166)
Belief × MA		−0.395* (0.205)
MA		1.266 (1.236)
Predicted contribution × TR		0.169 (0.164)
Belief × TR		−0.357* (0.194)
TR		−0.056 (1.100)
Socio-economic controls	Yes	Yes
Constant	−2.930 (2.223)	−2.488 (2.493)
R^2	0.59	0.60
No. obs.	360	360

Notes: OLS estimates with robust *SE* in parentheses. WEIRD cluster: UK and US; non-WEIRD cluster: Morocco and Turkey. Socio-economic control variables: age, female, urban background, middle class, single child, economics/business student. None is significant at any conventional level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

5.4. Discussion

Punishment in a one-shot game is a costly expression of a dislike of others' contributions that cannot be rationalized as a strategic investment because there are no future interactions. Nevertheless, our subjects did punish those group members who contributed less than them. However, surprisingly, antisocial punishment was slightly lower in non-WEIRD subject pools than in WEIRD subject pools, partially rejecting Hypothesis 2a. Yet, these differences in antisocial punishment between our WEIRD and non-WEIRD subject pools are rather small. Like in the situation without punishment, the biggest difference we observe is regarding beliefs about others' cooperativeness, where we confirm Hypothesis 2b of higher beliefs and higher cooperation in the WEIRD cluster than the non-WEIRD cluster.

Because preferences and beliefs differ (to some extent) between the WEIRD and non-WEIRD subject pools, an obvious question is what is more important – beliefs or preferences? We address this question by quantitatively assessing the relative importance of cooperative dispositions (a_i^s) and beliefs (b_i^s) to explain cross-cultural differences in contributions (c_i^s).

6. Explaining differences in cooperation: beliefs or preferences?

We conduct a simulation inspired by Fosgaard et al. (2014) to decompose the relative effects of a_i^s and b_i^s to explain cross-cultural differences in cooperation (c_i^s). Here, we describe the main steps of our analysis and Online Appendix I provides the details. To have enough power, we pool the data at the cluster level and compare those.

The first step consists of estimating the marginal effects of beliefs, predicted contributions and remaining cluster differences on contribution behaviour using a regression model, like that reported in Tables 3 and 4, on the full sample. Applying the conceptual framework described in Section 3.1, we assume a linear relationship between contributions and predicted contributions or beliefs.

The second step entails bootstrapping by randomly drawing (with replacement) 90 contribution tables (our measure of preferences a_i^c) and 90 beliefs b_i^c from each cluster $c \in \{WEIRD, non\text{-}WEIRD\}$ (90 corresponds to the average country sample size). Using the coefficient estimates from Step 1, we calculate four values of predicted contributions: Starting with a prediction based on the randomly drawn values from the WEIRD cluster, we sequentially replace beliefs, contribution tables, and the non-WEIRD dummy to the values reflecting the non-WEIRD cluster. The absolute differences between predicted contributions based on values from the WEIRD and non-WEIRD cluster represents the predicted total effect of cross-cluster differences on behaviour and can be decomposed in the belief effect, preference effect and the remaining unexplained effect captured by the dummy variable. In Step 2, we repeat the random draw 100 times to obtain bootstrapped standard errors for the predicted values.

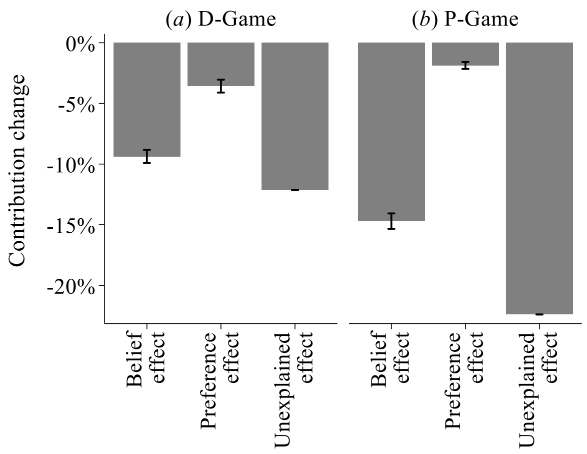


FIGURE 7. Decomposing the cross-cluster difference in contribution behaviour into the belief, preference and unexplained effect for the D-Game (*a*) and the P-Game (*b*). The negative effects reflect that the WEIRD cluster serves as the baseline. Percentage change relative to the sample average contribution in the respective game. The error bars indicate ±1 bootstrapped *SEM*.

For the D-Game (Figure 7a), we find a −25% difference in contributions (relative to the sample average) in the non-WEIRD compared to the WEIRD cluster. This total effect can be split into a −9% belief effect, −4% preference effect and −12% are unexplained. In the P-Game (Figure 7b), the predicted cluster differences in contributions amount to −39% (relative to the sample average) in the non-WEIRD compared to the WEIRD cluster, comprising a −15% belief effect, −2% preference effect with −22% remaining unexplained. Consistent with the evidence presented in the previous sections, beliefs emerge as the main factor explaining across-cluster differences in contributions in both games.

7. Concluding discussion

We presented a behavioural framework to study voluntary cooperation and applied it to investigate the cultural generalizability of two behavioural mechanisms that underpin voluntary cooperation: (1) cooperative dispositions and beliefs about others' cooperativeness, jointly explain cooperation; and (2) punishment. In public goods experiments conducted in two classically WEIRD societies, the UK, and the US and two non-WEIRD societies, Morocco and Turkey, we found that conditionally cooperative dispositions are similar across the four subject pools, and it is differences in beliefs that explain a significant share of the variation in unconditional cooperative behaviour. In a second experiment with peer punishment, we found stark differences in the participants' reaction to the punishment possibility. Only participants in the UK and the US significantly increased contributions in the game with punishment compared to the game without punishment. We found that actual punishment is similar across all four samples. Hence, social norms of cooperation and reactions to free riding are generalizable across our four samples.

Our results bear several insights for understanding human cooperation. First, conditional cooperation is a prevalent disposition in all four subject pools studied here and therefore not confined to WEIRD subject pools. Second, our results from the public goods experiments with and without punishment, as well as a simulation exercise, suggest that observed differences in cooperation across societies only stem to a small degree from variation in the distribution of cooperative dispositions across societies but are mostly driven by differences in beliefs about others' cooperativeness. Our results thus highlight the importance of beliefs about others' cooperativeness and how they might be a channel through which voluntary cooperation can be promoted across different societies. Although influencing beliefs is challenging, it is arguably more achievable than changing one's cooperative disposition, which likely is stable over longer periods of time (e.g., Bruhin et al. (2019); Chuang and Schechter (2015); Carlsson et al. (2014); Volk et al. (2012)) and thus appears more deeply ingrained through socialization.

Third, we found very little antisocial punishment (close to zero for Morocco and Turkey and slightly positive for UK and US), which is interesting because it is at odds with previous research reporting substantial antisocial punishment across societies in repeated games (Herrmann et al. (2008)) and one-shot games (Gächter and Herrmann (2009); Gächter and Herrmann (2011)). In repeated games with peer punishment, several mechanisms can explain the observed pattern of antisocial punishment. First, the punishment of high contributors who, however, fail to fully contribute to the public good might result from a strategic intention to maintain or increase the contributions of others. Second, antisocial punishment might result from the desire of a low contributor in a given round to pre-emptively retaliate against expected punishment from a higher contributor. Third, antisocial punishment might be an expression of inequality aversion (Thöni (2014)). Finally, antisocial punishment might be revenge for previous game play and thus take on the characteristics of a feud over several rounds (Nikiforakis and Engelmann (2011)). The one-shot design of the present study excludes any strategic or self-beneficial incentive to punish, does not allow for within-period retaliatory punishment (since there is only one punishment stage), and excludes the possibility of retaliation across periods. Therefore, we interpret punishment behaviour as an elicitation of a preference for sanctioning group members, mostly free riding behaviour. Our results show that there is no systematic variation in such preferences across the clusters we study.

In summary, as the four countries from which we drew our samples vary across many cultural and institutional dimensions, we can make a strong case for our two main findings: First, differences in beliefs about others' cooperativeness are more important than differences in cooperative dispositions to drive differences in cooperation between the subject pools we study here with the two WEIRD societies UK and US having higher beliefs than the non-WEIRD societies Morocco and Turkey. Second, punishment behaviour is very similar in all subject pools, which suggests that people have similar preferences for the enforcement of cooperation, regardless of the WEIRD-ness of the society they live in.

We conclude by noting that this paper provides evidence for the general applicability of a framework to study the behavioural mechanisms that underpin voluntary cooperation. Thus, it is not limited to the specific context in which we applied it here. Future research can therefore apply our framework to analyse the behavioural determinants of voluntary cooperation of any groups of interest.

References

Alesina, A., Giuliano, P., 2015. Culture and Institutions. Journal of Economic Literature 53, 898-944.

Algan, Y., Cahuc, P., 2013. Trust and Growth. Annual Review of Economics 5, 521-549.

Apicella, C., Norenzayan, A., Henrich, J., 2020. Beyond WEIRD: A review of the last decade and a look ahead to the global laboratory of the future. Evolution and Human Behavior 41, 319-329.

Balliet, D., Van Lange, P.A.M., 2013. Trust, Conflict, and Cooperation: A Meta-Analysis. Psychological Bulletin 139, 1090-1112.

Bigoni, M., Bortolotti, S., Casari, M., Gambetta, D., 2019. At the Root of the North-South Cooperation Gap in Italy: Preferences or Beliefs? Economic Journal 129, 1139-1152.

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

Bisin, A., Verdier, T., 2011. Chapter 9 - The Economics of Cultural Transmission and Socialization, in: Benhabib, J., Bisin, A., Jackson, M.O. (Eds.), Handbook of Social Economics. North-Holland, pp. 339-416.

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

Bluffstone, R., Dannenberg, A., Martinsson, P., Jha, P., Bista, R., 2020. Cooperative behavior and common pool resources: Experimental evidence from community forest user groups in Nepal. World Development 129, 104889.

Bohnet, I., Greig, F., Herrmann, B., Zeckhauser, R., 2008. Betrayal aversion. Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. Journal Article 98, 294-310.

Bohnet, I., Herrmann, B., Zeckhauser, R., 2010. Trust and the reference points for trustworthiness in Gulf and Western countries. Quarterly Journal of Economics 125, 811-828.

Brislin, R.W., 1970. Back-Translation for Cross-Cultural Research. Journal of Cross-Cultural Psychology 1, 185-216.

Brosig-Koch, J., Helbach, C., Ockenfels, A., Weimann, J., 2011. Still different after all these years: Solidarity behavior in East and West Germany. Journal of Public Economics 95, 1373-1376.
Bruhin, A., Fehr, E., Schunk, D., 2019. The many Faces of Human Sociality: Uncovering the Distribution and Stability of Social Preferences. Journal of the European Economic Association 17, 1025-1069.
Bruhin, A., Janizzi, K., Thöni, C., 2020. Uncovering the heterogeneity behind cross-cultural variation in antisocial punishment. Journal of Economic Behavior & Organization 180, 291-308.
Carlsson, F., Johansson-Stenman, O., Nam, P.K., 2014. Social preferences are stable over long periods of time. Journal of Public Economics 117, 104-114.
Carpenter, J.P., Matthews, P.H., 2012. Norm Enforcement: Anger, Indignation, or Reciprocity? Journal of the European Economic Association 10, 555-572.
Cavalli-Sforza, L., Menozzi, P., Piazza, A., 1994. The history and geography of human genes. Princeton University Press, Princeton NJ.
Chaudhuri, A., 2011. Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. Experimental Economics 14, 47-83.
Choi, S., Kim, B.-Y., Lee, J., Lee, S., 2020. A tale of two Koreas: Property rights and fairness. Journal of Economic Behavior & Organization 170, 112-130.
Chuah, S.H., Gächter, S., Hoffmann, R., Tan, J.H.W., 2016. Religion, discrimination and trust across three cultures. European Economic Review 90, 280-301.
Chuang, Y., Schechter, L., 2015. Stability of experimental and survey measures of risk, time, and social preferences: A review and some new results. Journal of Development Economics 117, 151-170.
Cohn, A., Maréchal, M.A., Tannenbaum, D., Zünd, C.L., 2019. Civic honesty around the globe. Science 365, 70-73.
Cubitt, R., Drouvelis, M., Gächter, S., 2011. Framing and free riding: emotional responses and punishment in social dilemma games. Experimental Economics 14, 254–272.
Cubitt, R., Gächter, S., Quercia, S., 2017. Conditional cooperation and betrayal aversion. Journal of Economic Behavior & Organization 141, 110-121.
Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., Sunde, U., 2018. Global Evidence on Economic Preferences. Quarterly Journal of Economics 133, 1645–1692.
Fehr, E., 2009. The economics and biology of trust. Journal of the European Economic Association 7, 235-266.
Fehr, E., Gächter, S., 2002. Altruistic punishment in humans. Nature 415, 137-140.
Fehr, E., Schurtenberger, I., 2018. Normative foundations of human cooperation. Nature Human Behaviour 2, 458-468.
Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Exp Econ 10, 171-178.
Fischbacher, U., Gächter, S., 2010. Social preferences, beliefs, and the dynamics of free riding in public good experiments. *American Economic Review.* 100, 541–556.
Fischbacher, U., Gächter, S., Fehr, E., 2001. Are people conditionally cooperative? Evidence from a public goods experiment. Econ Lett 71, 397-404.
Fischbacher, U., Gächter, S., Quercia, S., 2012. The behavioral validity of the strategy method in public good experiments. Journal of Economic Psychology 33, 897-913.
Fisman, R., Miguel, E., 2007. Corruption, Norms, and Legal Enforcement: Evidence from Diplomatic Parking Tickets. The Journal of Political Economy 115, 1020-1048.
Fosgaard, T.R., Hansen, L.G., Wengström, E., 2014. Understanding the nature of cooperation variability. Journal of Public Economics 120, 134-143.
Frey, B.S., Meier, S., 2004. Social comparisons and pro-social behavior. Testing 'conditional cooperation' in a field experiment. American Economic Review. 94, 1717-1722.
Gächter, S., 2007. Conditional cooperation: Behavioral regularities from the lab and the field and their policy implications, in: Frey, B.S., Stutzer, A. (Eds.), Psychology and Economics: A Promising New Cross-Disciplinary Field (CESifo Seminar Series). The MIT Press, Cambridge, pp. 19-50.
Gächter, S., 2010. (Dis)advantages of student subjects: What is your research question? Behavioral and Brain Sciences 33, 92-93.

Gächter, S., Herrmann, B., 2009. Reciprocity, culture, and human cooperation: previous insights and a new cross-cultural experiment. Philosophical Transactions of the Royal Society B – Biological Sciences 364, 791-806.
Gächter, S., Herrmann, B., 2011. The limits of self-governance when cooperators get punished: Experimental evidence from urban and rural Russia. European Economic Review 55, 193-210.
Gächter, S., Herrmann, B., Thöni, C., 2004. Trust, voluntary cooperation, and socio-economic background: survey and experimental evidence. Journal of Economic Behavior & Organization 55, 505-531.
Gächter, S., Herrmann, B., Thöni, C., 2010. Culture and cooperation. Philosophical Transactions of the Royal Society B: Biological Sciences 365, 2651-2661.
Gächter, S., Kölle, F., Quercia, S., 2017. Reciprocity and the tragedies of maintaining and providing the commons. Nature Human Behaviour 1, 650-656.
Gächter, S., Renner, E., 2010. The effects of (incentivized) belief elicitation in public goods experiments. Experimental Economics 13, 364-377.
Gächter, S., Renner, E., 2018. Leaders as role models and 'belief managers' in social dilemmas. Journal of Economic Behavior & Organization 154, 321-334.
Gächter, S., Schulz, J.F., 2016. Intrinsic honesty and the prevalence of rule violations across societies. Nature 531, 496-499.
Greif, A., 1994. Cultural beliefs and the organization of society: A historical reflection on collectivist and individualist societies. Journal of Political Economy 102, 912-950.
Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. J Econ Sci Assoc 1, 114-125.
Guiso, L., Sapienza, P., Zingales, L., 2006. Does culture affect economic outcomes? Journal of Economic Perspectives 20, 23-48.
Guiso, L., Sapienza, P., Zingales, L., 2008. Social Capital as Good Culture. Journal of the European Economic Association 6, 295-320.
Hauk, E., Saez-Marti, M., 2002. On the Cultural Transmission of Corruption. Journal of Economic Theory 107, 311-335.
Henrich, J., 2015. The Secret of Our Success: How Culture is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter. Princeton University Press, Princeton.
Henrich, J., 2020. The WEIRDest People in the World: How the West Became Psychologically Peculiar and Particularly Prosperous. Farrar, Strauss and Giroux, New York.
Henrich, J., Boyd, R., Bowles, S., Camerer, C.F., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Henrich, N., Hill, K., Gil-White, F., Gurven, M., Marlowe, F.W., Patton, J.Q., Tracer, D., 2005. "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. Behavioral and Brain Sciences 28, 795-855.
Henrich, J., Heine, S.J., Norenzayan, A., 2010. The weirdest people in the world? Behavioral and Brain Sciences 33, 61-83.
Herrmann, B., Thöni, C., 2009. Measuring conditional cooperation: a replication study in Russia. Experimental Economics 12, 87-92.
Herrmann, B., Thöni, C., Gächter, S., 2008. Antisocial punishment across societies. Science 319, 1362-1367.
Hofstede, G.H., Hofstede, G., 2001. Culture's consequences: Comparing values, behaviors, institutions and organizations across nations. Sage Publications, Thousand Oaks, CA.
Inglehart, R., Haerpfer C., Moreno A., Welzel C., Kizilova K., Diez-Medrano J., Lagos M., Norris P., Ponarin E., (eds.), P.B.e.a., 2014. World Values Survey: Round Five - Country-Pooled Datafile Version: www.worldvaluessurvey.org/WVSDocumentationWV5.jsp., Madrid: JD Systems Institute.
Inglehart, R., Welzel, C., 2005. Modernization, Cultural Change, and Democracy. The Human Development Sequence. Cambridge University Press, Cambridge.
Kaufmann, D., Kraay, A., 2016. Worldwide Governance Indicators, Dataset retrieved from: http://www.govindicators.org.
Keser, C., van Winden, F., 2000. Conditional cooperation and voluntary contributions to public goods. Scandinavian Journal of Economics 102, 23-39.
Kim, B.-Y., Choi, S., Lee, J., Lee, S., Choi, K., 2017. Do Institutions Affect Social Preferences? Evidence from Divided Korea. Journal of Comparative Economics 45, 865-888.

Knack, S., Keefer, P., 1997. Does Social Capital Have an Economic Payoff? A Cross-Country Investigation. Quarterly Journal of Economics 112, 1251-1288.
Kocher, M.G., Cherry, T., Kroll, S., Netzer, R.J., Sutter, M., 2008. Conditional cooperation on three continents. Economics Letters 101, 175-178.
Kocher, M.G., Martinsson, P., Matzat, D., Wollbrant, C., 2015. The role of beliefs, trust, and risk in contributions to a public good. Journal of Economic Psychology 51, 236-244.
Martinsson, P., Pham-Khanh, N., Villegas-Palacio, C., 2013. Conditional cooperation and disclosure in developing countries. Journal of Economic Psychology 34, 148-155.
Molleman, L., Gächter, S., 2018. Societal background influences social learning in cooperative decision making. Evolution and Human Behavior 39, 547-555.
Muthukrishna, M., Bell, A.V., Henrich, J., Curtin, C.M., Gedranovich, A., McInerney, J., Thue, B., 2020. Beyond Western, Educated, Industrial, Rich, and Democratic (WEIRD) Psychology: Measuring and Mapping Scales of Cultural and Psychological Distance. Psychological Science 31, 678-701.
Nikiforakis, N., 2008. Punishment and counter-punishment in public good games: Can we really govern ourselves? Journal of Public Economics 92, 91-112.
Nikiforakis, N., Engelmann, D., 2011. Altruistic punishment and the threat of feuds. Journal of Economic Behavior & Organization 78, 319-332.
Oosterbeek, H., Sloof, R., van de Kuilen, G., 2004. Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. Experimental Economics 7, 171-188.
Prediger, S., Vollan, B., Herrmann, B., 2014. Resource scarcity and antisocial behavior. Journal of Public Economics 119, 1-9.
Romano, A., Balliet, D., Yamagishi, T., Liu, J.H., 2017. Parochial trust and cooperation across 17 societies. Proceedings of the National Academy of Sciences 114, 12702-12707.
Roth, A.E., Prasnikar, V., Okuno-Fujiwara, M., Zamir, S., 1991. Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: an experimental study. American Economic Review. 81, 1068-1095.
Rustagi, D., 2020. Historical Self-Governance and Norms of Cooperation. Working Paper, Brown University No.
Rustagi, D., Engel, S., Kosfeld, M., 2010. Conditional Cooperation and Costly Monitoring Explain Success in Forest Commons Management. Science 330, 961-965.
Schulz, J., 2020. Kin Networks and Institutional Development. Manuscript, George Mason University No.
Schulz, J.F., Bahrami-Rad, D., Beauchamp, J.P., Henrich, J., 2019. The Church, intensive kinship, and global psychological variation. Science 366, eaau5141.
Sylwester, K., Herrmann, B., Bryson, J.J., 2013. *Homo homini lupus*? Explaining antisocial punishment. Journal of Neuroscience, Psychology, and Economics 6, 167-188.
Tabellini, G., 2008. The scope of cooperation: Values and incentives. Quarterly Journal of Economics 123, 905-950.
Thöni, C., 2014. Inequality aversion and antisocial punishment. Theory and Decision 76, 529-545.
Thöni, C., 2019. Cross-cultural behavioral experiments: potential and challenges, in: Schram, A., Ule, A. (Eds.), Handbook of Research Methods and Applications in Experimental Economics. Edward Elgar Publishing, Cheltenham.
Thöni, C., Tyran, J.-R., Wengström, E., 2012. Microfoundations of social capital. Journal of Public Economics 96, 635-643.
Thöni, C., Volk, S., 2018. Conditional Cooperation: Review and Refinement. Economics Letters 171, 37-40.
Volk, S., Thöni, C., Ruigrok, W., 2012. Temporal stability and psychological foundations of cooperation preferences. Journal of Economic Behavior & Organization 81, 664-676.
Weber, T.O., Weisel, O., Gächter, S., 2018. Dispositional free riders do not free ride on punishment. Nature Communications 9, 2390.

ONLINE APPENDIX

The Behavioural Mechanisms of Voluntary Cooperation in WEIRD and Non-WEIRD Societies

Till O. Weber[1,†], Benjamin Beranek[2], Simon Gächter[3,4,5], Fatima Lambarraa-Lehnhardt[6,7], and Jonathan F. Schulz[8]

[1] Newcastle University Business School, Newcastle University, UK; tilloweber@gmail.com

[2] Department of Economics, Missouri State University, USA; benjaminberanek@missouristate.edu

[3] School of Economics, University of Nottingham, UK; simon.gaechter@nottingham.ac.uk

[4] IZA Bonn, Germany; [5] CESifo Munich, Germany

[6] Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg, Germany; fatima.lehnhardt@zalf.de [7] CRC-PEG, University of Göttingen, Germany

[8] Department of Economics, George Mason University, USA, jonathan.schulz77@gmail.com

[†] Corresponding author

June 12, 2021

A. Subject pool differences **2**

B. Alternative classifications of cooperative dispositions **5**

C. Conditional contributions across subject pools **6**

D. Accuracy of beliefs **8**

E. Accuracy of predicted contributions **8**

F. Testing the independence of dispositions and punishment **9**

G. Regression analysis of expected and realized punishment **10**

H. Accuracy of expected punishment **15**

I. Assessing the relative importance of beliefs and preferences for cooperation **16**

J. Instructions for the D-Game **19**

K. Instructions for the P-Game **25**

L. Decision screens not included in the instructions **28**

M. Socio-economic background questionnaire **30**

References **31**

A. Subject pool differences

Institutional and cultural differences across the four subject pools are likely to be captured by various macro-level indicators (Herrmann, Thöni, and Gächter 2008), summarized in Table A1.

TABLE A.1. Cultural and institutional indicators for Morocco, Turkey, the UK and the US.

Dimension	Indicator	US	UK	MA	TR
Cultural dimensions and world values	Traditional vs. secular-rational	−0.81	0.06	−1.32	−0.89
	Survival vs. self-expression	1.76	1.68	−1.04	−0.33
	Power distance	40	35	70	66
	Individualism	91	89	46	37
Institutional quality	Government effectiveness	1.46	1.62	−0.14	0.38
	Prevalence of Rule Violations	−2.78	−2.93	0.47	0.04
	GDP per capita	53.01	38.84	7.30	23.39
	Rule of law	1.62	1.89	−0.06	0.04
Social capital and preferences	Norms of civic cooperation	8.81	8.72	9.07	9.40
	Perceived trust	39.56	30.43	13.00	4.78
	Positive reciprocity	0.20	−0.04	0.55	−0.44
	Negative reciprocity	0.01	0.02	−0.49	0.52

Notes: *Traditional vs. Secular-Rational Values* (−1.94 = strongest emphasis on traditional values; 1.96 = strongest emphasis on secular-rational values) and *Survival vs. Self-Expression Values* (−1.68 = strongest emphasis on survival values; 2.35 = strongest emphasis on self-expression values) from Inglehart and Welzel (2005); *Power Distance* (11 = lowest; 104 = highest) and *Individualism* (6 = least individualist; 91 = most individualist) from Hofstede and Hofstede (2001); *Government Effectiveness* (−2.48 = lowest government effectiveness; 2.19 = highest government effectiveness) from the 2014 Worldwide Governance Indicators (Kaufmann and Kraay (2016)); *Prevalence of Rule Violations* (−3.10 = lowest; 2.84 = highest) from Gächter and Schulz (2016); World Bank *GDP Per Capita* in current USD (PPP) for 2015; *Rule of Law* (−2.39 = lowest rule of law; 2.12 = highest rule of law) from the 2014 Worldwide Governance Indicators (Kaufmann and Kraay (2016)); *Norms of Civic Cooperation* (1 = very weak norm of civic cooperation; 10 = very strong norm of civic cooperation) based on data from the World Values Survey Wave 5 (Knack and Keefer (1997)); *Perceived Trust*: percent of people who agree with the statement “most people can be trusted” by Inglehart et al. (2014); *Positive Reciprocity* (−1.04 = lowest positive reciprocity; 0.57 = highest positive reciprocity) and *Negative Reciprocity* (−0.49 = lowest negative reciprocity; 0.74 = highest negative reciprocity) from the Global Preference Survey (Falk et al. (2018)).

Inglehart and Welzel (2005) measure cultural values using responses from the World Values Survey (WVS). They argue that societal differences can be measured using two dimensions: *traditional values versus secular-rational values* and *survival values versus self-expression values*. The first dimension measures the importance of authority, traditional family values and religion in a society. The second dimension indicates the valuation of self-expression, individual wellbeing and quality of life. The two dimensions are extracted from the WVS responses using factor analysis and together account for 71% of the cross-national variation (Inglehart and Welzel 2005, p. 50). Both, the US and the UK score high on self-expression values, but the importance of traditional and secular-rational values differs. The US scores higher on traditional values whereas the UK scores higher on secular-rational values. Morocco and Turkey score below the sample average in both dimensions showing a high emphasis on traditional values and survival values.

Hofstede's cultural dimensions (Hofstede and Hofstede 2001) are a long-established set of dimensions to quantify cultural differences: *Individualism* measures the importance of the collective versus the individual. The degree of individualism or collectivism is a fundamental characteristic of any society (Greif 1994). *Power distance* measures how unequal or egalitarian a society is. For the US and UK, individualism scores are amongst the highest of the sample and they score below the sample average for power distance. Both, Morocco and Turkey, score above the sample average for power distance. For individualism, Turkey is just below and Morocco slightly above the sample average.

Amongst the indicators of institutional quality are *Government effectiveness* and *rule of law* are drawn from the Worldwide Governance Indicators 2015 (Kaufmann, Kraay, and Mastruzzi 2011; Kaufmann and Kraay 2016) and measure institutional quality. The US and UK score higher than the sample average for government effectiveness and rule of law. Morocco and Turkey score close to the sample average for government effectiveness and below the average for rule of law. These measures are directly linked to the quality of formal institutions such as the government and the judicial system. The *GDP per capita* (PPP) of both, the US and UK, is far above the sample average, whereas that of Morocco and Turkey lies below the sample average. The *prevalence of rule violations* (PRV, Gächter and Schulz 2016) measures how common rule violations like corruption, tax evasion or political fraud are in a society. PRV is close to the sample average for the US and UK but lies below the average for Morocco and Turkey.

In addition to measures of culture and institutional quality, we draw on indicators of social capital and social preferences to measure the distance between societies. *Norms of civic cooperation* and *perceived trust* relate to the strength of social norms in a society and are taken from the WVS. We calculate a score for norms of civic cooperation based on data from Wave 5 of the World Values survey (Knack and Keefer 1997). The scores measure the acceptability of claiming government benefits one is not entitled to, fare-dodging on public transport and cheating on taxes (rescaled average value of WVS items V198-V200). All four countries' scores for norms of civic cooperation are higher than the sample average. Perceived trust in the US and UK lie above the sample average, whereas the scores for Morocco and Turkey are lower than the world average. The propensity to act in a positively and negatively reciprocal way might be an important factor of cooperation success in a society. We report *positive reciprocity* and *negative reciprocity* scores from the Global Preference Survey (Falk et al. 2018). We find a large societal variation in positive and negative reciprocity across the four societies with no clear pattern across the WEIRD and non-WEIRD cluster.

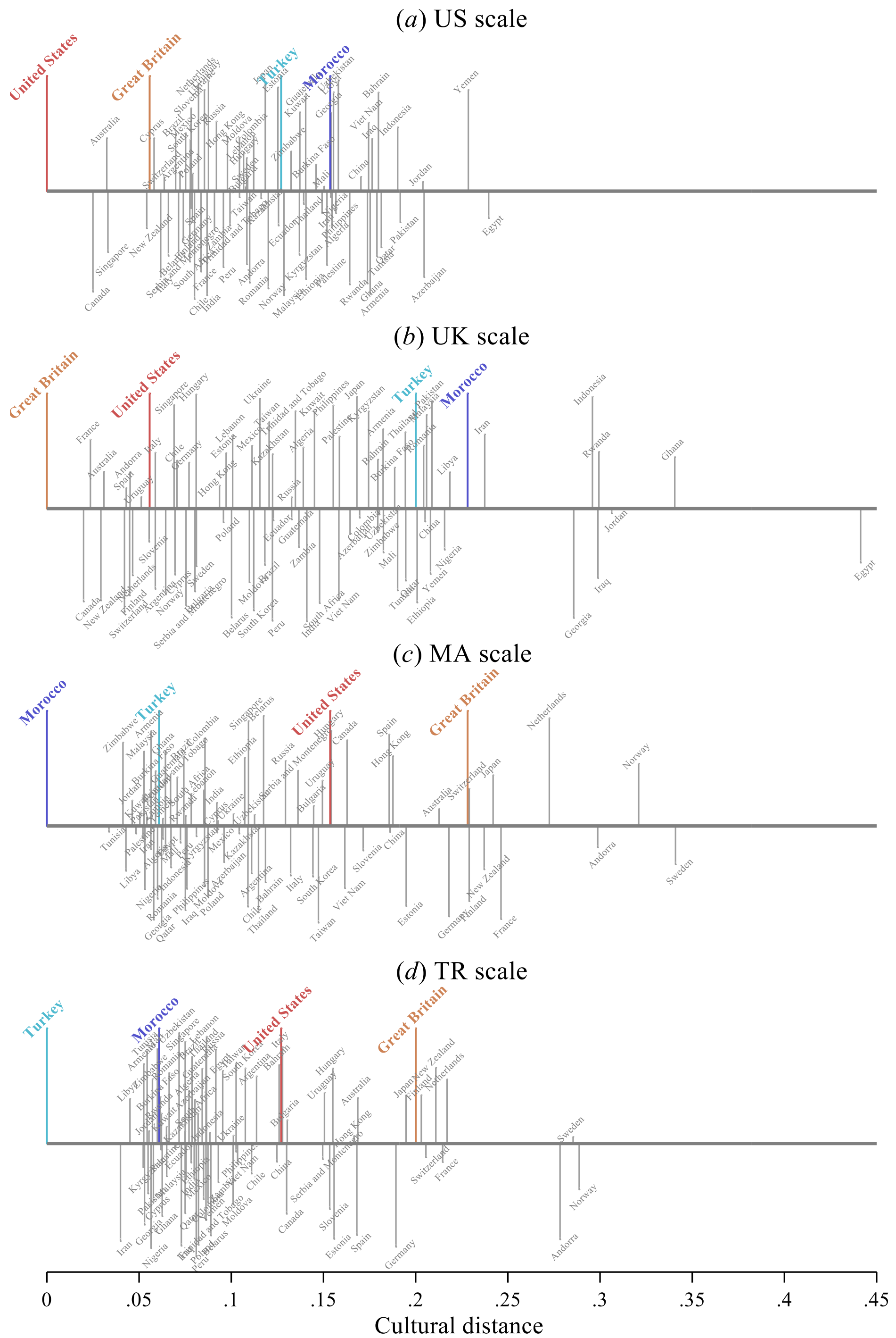


FIGURE A.1. Cultural distance from the respective benchmark country as calculated by Muthukrishna et al. (2020).

The cultural distance measures as calculated by Muthukrishna et al. (2020) are illustrated in Figure A1. Calculation details are in Footnotes 7 and 8 of the main text. The data are taken from http://culturaldistance.com. Panel (*a*) illustrates cultural distance from the US; panel (*b*) from the UK; panel (*c*) from Morocco and panel (*d*) from Turkey.

B. Alternative classifications of cooperative dispositions

To check the robustness of our findings, we compare the distribution of cooperative dispositions across the four samples using the original classification of types by Fischbacher, Gächter, and Fehr (2001). Again, we find that CC account for the largest share of participants in all four samples with no significant difference in the share of CC across the four subject pools ($\chi^2(3) = 2.97$, $p = 0.396$). The shares of FR ($\chi^2(3) = 9.19$, $p = 0.027$) and the shares of OT ($\chi^2(3) = 7.63$, $p = 0.054$) vary across the four samples.

The distributions of cooperative dispositions are statistically significantly different across the four samples ($\chi^2(6) = 14.33$, $p = 0.026$). Comparing the WEIRD and the non-WEIRD cluster reveals weakly significant differences in the distributions of cooperative dispositions. Within clusters, there are no significant differences when comparing the distributions between the US and UK samples and weakly significant differences in the distribution of cooperative dispositions between the MA and TR samples.

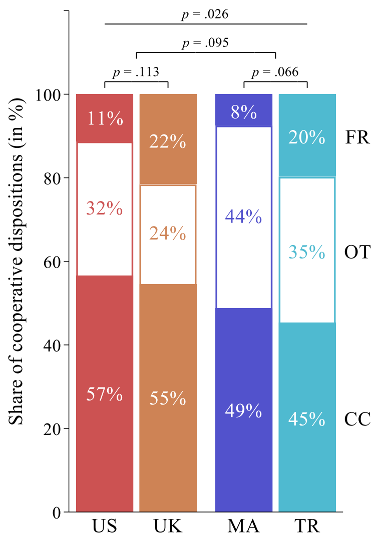


FIGURE B.1. The share of conditional cooperators (CC), free riders (FR) and others (OT) according to the classification by Fischbacher et al (2001); *p*-values from a (pooled) χ^2 tests.

C. Conditional contributions across subject pools

TABLE C1. Conditional contributions of CC across subject pools.

Dependent variable: conditional contribution	(1) By cluster	(2) By country
Average contributions of others	0.801*** (0.029)	0.802*** (0.037)
Average contribution of others × Non-WEIRD cluster	−0.193*** (0.048)	
Non-WEIRD cluster	0.242 (0.482)	
Average contribution of others × UK		−0.004 (0.059)
UK		0.728 (0.602)
Average contribution of others × MA		−0.284*** (0.062)
MA		0.892 (0.541)
Average contribution of others × TR		−0.109* (0.066)
TR		0.365 (0.712)
Socio-economic controls	Yes	
Constant	4.181 (2.532)	4.542* (2.618)
R^2	0.52	0.53
No. obs. (Clusters)	3822 (182)	3822 (182)

Notes: Includes only CC. OLS estimation with robust *SE* clustered on individuals in parentheses. The dependent variable is the conditional contribution elicited on the contribution table in the D-Game where participants indicated their own conditional contribution for each of the 21 average contributions of their group members. Socio-economic control variables: age, female, urban background, middle class, single child, economics/business student. In Col. 1, only child is negative and weakly significant. In Col. 2, age is negative and weakly significant. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table C1 Column 2 shows a statistically insignificant interaction term "Average contribution of others × UK" and UK dummy, indicating no within-cluster variation for the WEIRD cluster. We find some within-cluster variation for the non-WEIRD cluster with a lower increase of conditional contributions for a higher average contribution of others in the MA sample ($F(1, 181) = 5.71$, $p = 0.018$) but similar levels of conditional contributions in the MA and TR samples ($F(1, 181) = 0.62$, $p = 0.431$).

TABLE C2. Conditional contributions of OT across subject pools.

Dependent variable: conditional contribution	(1) By cluster	(2) By country
Average contributions of others	0.028 (0.034)	0.049 (0.039)
Average contribution of others × Non-WEIRD cluster	−0.029 (0.039)	
Non-WEIRD cluster	0.027 (0.999)	
Average contribution of others × UK		−0.054 (0.073)
UK		2.707** (1.333)
Average contribution of others × MA		−0.064 (0.050)
MA		0.304 (1.133)
Average contribution of others × TR		−0.032 (0.042)
TR		2.010 (1.397)
Socio-economic controls	Yes	Yes
Constant	−2.044 (6.226)	−2.946 (6.120)
R^2	0.03	0.06
No. obs. (Clusters)	2604 (124)	2604 (124)

Notes: Includes only OT. OLS estimation with robust *SE* clustered on individuals in parentheses. The dependent variable is the conditional contribution elicited on the contribution table in the D-Game where participants indicated their own conditional contribution for each of the 21 average contributions of their group members. Socio-economic control variables: age, female, urban background, middle class, single child, economics/business student. None is significant at any conventional level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table C2 Column 2 shows a statistically insignificant interaction term “Average contribution of others × UK” but a positive and significant UK dummy, indicating a higher level of conditional contributions in the UK compared to the US sample. For the non-WEIRD cluster, we find no evidence for within-cluster variation with a statistically similar increase of conditional contributions for a higher average contribution of others ($F(1, 123) = 0.79$, $p = 0.376$) and statistically similar levels of conditional contributions in the MA and TR samples ($F(1, 123) = 1.03$, $p = 0.311$).

D. Accuracy of beliefs

Across the four samples, participants are on average imperfect conditional cooperators who condition their own kindness on their expected behaviour of others. This conditional cooperation strategy works best if people can predict the behaviour of others correctly. Furthermore, the accuracy of beliefs might be an important indicator for the saliency of contribution norms within a society. We define belief accuracy as the deviation of beliefs about other group members' behaviour from their actual unconditional contribution (Table D.1). In the D-Game, we find a similar accuracy of beliefs across samples (Kruskal-Wallis tests, $\chi^2(3) = 4.31$, $p = 0.230$). Thus, for this game, there is some uncertainty about other people's behaviour, but this uncertainty does not vary across the WEIRD and non-WEIRD cluster. In the P-Game We find significant differences in belief accuracy across samples (Kruskal-Wallis tests, $\chi^2(3) = 8.93$, $p = 0.030$). These differences stem from significantly different belief accuracies across clusters (pooled Mann-Whitney $Z = -2.46$, $p = 0.014$). We find no significant within-cluster differences in belief accuracy for the WEIRD cluster (Mann-Whitney $Z = 1.52$, $p = 0.129$) or the non-WEIRD cluster (Mann-Whitney $Z = -1.02$, $p = 0.308$).

TABLE D.1. Deviation of beliefs from actual contributions (i.e., belief accuracy).

	D-Game	P-Game
United States (*N* = 106)	2.17 (4.88)	1.00 (5.38)
United Kingdom (*N* = 88)	0.79 (6.34)	−0.22 (6.16)
Morocco (*N* = 80)	1.60 (6.45)	1.59 (6.67)
Turkey (*N* = 86)	2.64 (6.41)	2.63 (6.74)

Notes: A positive value indicates that on average beliefs exceed contributions. *SD* in parentheses.

E. Accuracy of predicted contributions

Following Fischbacher, Gächter, and Quercia (2012), we calculate the predicted unconditional contribution for each participant using their schedule and unconditional belief. We then calculate the deviation between the predicted and the actual contribution (Figure E.1). 61% in the US, 61% in the UK, 61% in the MA and 63% in the TR sample are consistent with predicted contributions, allowing for a deviation of ±2 tokens. The average deviation from predicted contributions is not significantly different across the four samples (Kruskal-Wallis test, $\chi^2(3) = 1.14$, $p = 0.768$).

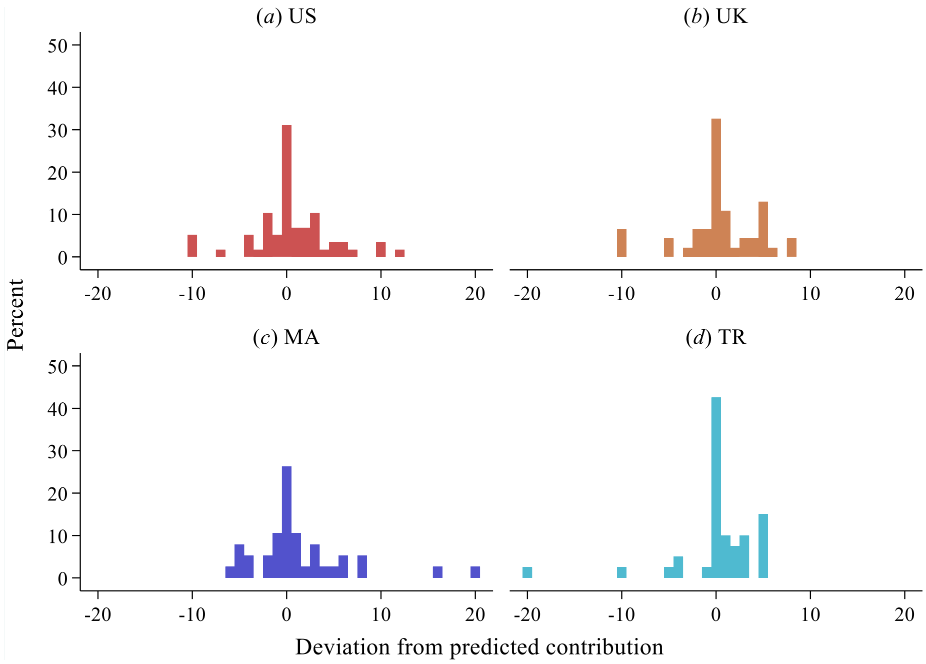


FIGURE E.1. Deviations from predicted contributions in the four countries. A positive value indicates that the actual contributions exceed predicted contributions.

F. Testing the independence of dispositions and punishment

Here we test the assumption that an individual's preference for punishment of others is independent from their cooperative disposition (that is, p_{ij}^s is independent of a_i^s – see Section 3.1 in the main paper; Weber, Weisel, and Gächter (2018); and Molleman et al. (2019). We find similar prosocial punishment expenditures when comparing the two most prevalent types of cooperative dispositions in the WEIRD cluster (pooling US & UK; $M_{CC} = 0.57$, $M_{FR} = 0.79$; Mann-Whitney $Z = 0.42$, $p = 0.690$; each participant as an independent observation) and non-WEIRD cluster (pooling Morocco & Turkey; $M_{CC} = 0.78$, $M_{FR} = 0.36$; Mann-Whitney $Z = 0.57$, $p = 0.695$; each participant as an independent observation). Furthermore, similar proportions of the two types engage in prosocial punishment in the WEIRD cluster (41% CC vs. 31% FR; $\chi^2(1) = 0.51$, $p = 0.477$) and non-WEIRD cluster (34% CC vs. 20% FR; $\chi^2(1) = 0.40$, $p = 0.525$). We therefore conclude that the assumption that p_i^s is independent of a_i^s is justified for the WEIRD and non-WEIRD samples included in this study. This means that in all samples studied here, free riders and conditional cooperators are equally likely to punish those who free ride.

G. Regression analysis of expected and realized punishment

We use Tobit regression models to test for differences in expected punishment across the WEIRD and non-WEIRD cluster (Table G.1). Although some authors suggest a two-stage regression model in order to separate between the likelihood of punishment and the punishment severity (Nikiforakis and Engelmann 2011; Weber, Weisel, and Gächter 2018), splitting the regression analysis in two steps is not feasible due to a small sample size and relatively few punishment incidences in our one-shot game.

As independent variables, we include the absolute negative deviation in the contribution levels between the punisher and the person receiving the punishment. Additionally, we include the positive contribution deviation between the punisher and the person receiving the punishment. We first estimate the models separately for the four societies (Table G.1; Col. 1-4). In all but the MA sample, subjects expect a significant increase in punishment for larger negative deviation from the punisher's contribution. In the US and TR samples, subjects' expected punishment also increases in the positive deviation from the punisher's contribution.

We also estimate a pooled model with further explanatory variables to test for differences between cultural clusters (Table G.1; Col. 5). We include a dummy variable for the non-WEIRD cluster and interaction terms between the dummy variable and the contribution deviation. Additionally, we include socio-economic controls (age, gender, urban background, middle class, single child, economics or business student). We find weakly significantly lower levels of expected punishment in the non-WEIRD compared to the WEIRD cluster ($b = -0.691$, $p = 0.089$). The highly significant negative interaction term "Absolute negative deviation × Non-WEIRD cluster" indicates a lower expected punishment for negative contribution deviations in the non-WEIRD cluster ($b = -0.131$, $p = 0.009$). The statistically insignificant interaction term "Positive negative deviation × Non-WEIRD cluster" suggests similar reactions to positive contribution deviations across clusters ($b = 0.015$, $p = 0.812$).

To test for within-cluster differences, we run a similar pooled regression model with UK, MA and TR dummies and the respective interaction terms (Table G.1; Col. 6). The weakly significant negative interaction term "Positive deviation × UK" suggests lower expected punishment for positive contribution deviations in the UK compared to the US sample ($b = -0.149$, $p = 0.070$). We find no evidence for differences in the expected punishment for negative contribution deviations ($b = -0.001$, $p = 0.988$) or level differences in expected punishment within the WEIRD cluster ($b = 0.775$, $p = 0.118$).

Next, we test for within-cluster differences in the non-WEIRD cluster by testing for significant differences in the coefficient sizes of the MA and TR dummies as well as their respective interaction terms. We find a significantly lower expected punishment for negative contribution deviations in the MA compared to the TR sample ($F(1, 1063) = 5.05$, $p = 0.025$), but a statistically similar expected punishment for positive contribution levels ($F(1, 1063) = 0.64$, $p = 0.426$) and expected punishment levels ($F(1, 1063) = 0.72$, $p = 0.396$).

To investigate societal differences in realized punishment, we estimate Tobit models with a similar specification as discussed above (Table G.2). Now, the dependent variable is the number of actual punishment points. The absolute negative deviation now refers to the deviation of a group member from the punisher and shows prosocial punishment. A positive deviation indicates higher contributions compared to the punisher which implies antisocial punishment.

Table G.1. Regression analysis of *expected* punishment.

Dependent variable: expected punishment	(1) US	(2) UK	(3) MA	(4) TR	(5) By cluster	(6) By country
Absolute negative deviation	0.205*** (0.048)	0.205*** (0.044)	−0.002 (0.081)	0.195*** (0.059)	0.240*** (0.035)	0.235*** (0.051)
Positive deviation	0.087 (0.056)	−0.043 (0.051)	−0.001 (0.094)	0.072 (0.060)	0.027 (0.042)	0.100* (0.059)
Absolute negative deviation × Non-WEIRD cluster					−0.131*** (0.050)	
Positive deviation × Non-WEIRD cluster					0.015 (0.062)	
Non-WEIRD cluster					−0.691* (0.406)	
Absolute negative deviation × UK						−0.001 (0.066)
Positive deviation × UK						−0.149* (0.082)
UK						0.775 (0.495)
Absolute negative deviation × MA						−0.238*** (0.082)
Positive deviation × MA						−0.108 (0.095)
MA						−0.535 (0.650)
Absolute negative deviation × TR						−0.057 (0.069)
Positive deviation × TR						−0.034 (0.081)
TR						0.081 (0.538)
Socio-economic control variables	No	No	No	No	Yes	Yes
Constant	−0.859** (0.391)	−0.360 (0.337)	−2.492*** (0.793)	−1.473*** (0.509)	−2.061 (1.385)	−2.079 (1.407)
Pseudo R^2	0.03	0.06	< 0.01	0.02	0.04	0.05
No. obs. (Clusters)	318 (106)	264 (88)	240 (80)	258 (86)	1080 (360)	1080 (360)

Notes: Tobit coefficients. Robust *SE* clustered on individuals. Socio-economic control variables: age, female, urban background, middle class, single child, economics/business student. In Col. 5, middle class is negative and significant, economist is negative and weakly significant. In Col. 6, female is positive and weakly significant, economist is negative and significant. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

First, we estimate regression analyses for each society separately (Table G.2; Col. 1-4). For all but the MA sample, the punishment expenditure increases significantly for a larger negative contribution deviation. For the TR sample, punishment decreases significantly if the other group member contributes more than the punisher.

Comparing differences across the WEIRD and non-WEIRD clusters yields no significant level differences in punishment expenditure ($b = 0.048$, $p = 0.944$; Table G.2; Col. 5). Yet, the negative and weakly significant interaction term "Absolute negative deviation × Non-WEIRD cluster" ($b = -0.171$, $p = 0.070$) and the negative and significant interaction term "Positive deviation × Non-WEIRD cluster" ($b = -0.276$, $p = 0.015$) indicate lower punishment of negative and positive deviations in the non-WEIRD compared to the WEIRD cluster. This shows that there is less punishment of free riding and less antisocial punishment in the non-WEIRD cluster.

To investigate differences in punishment expenditure within clusters, we estimate a similar regression model with UK, MA and TR dummies and their respective interaction terms (Table G.2; Col. 5). The statistically insignificant UK dummy and interaction terms "Absolute negative deviation × UK" ($b = 0.069$, $p = 0.603$) and "Positive deviation × UK" ($b = 0.087$, $p = 0.486$) reveal no variation in punishment expenditure within the WEIRD cluster.

Next, we test for within-cluster differences in the non-WEIRD cluster by testing for significant differences in the coefficient sizes of the MA and TR dummies as well as their respective interaction terms. We find no evidence for differences in punishment of negative contribution deviations ($F(1, 1063) = 2.65$, $p = 0.104$), positive contributions ($F(1, 1063) = 0.01$, $p = 0.926$) or level differences in punishment expenditure ($F(1, 1063) = 0.23$, $p = 0.629$).

Table G.2. Regression analysis of *actual* punishment.

Dependent variable: expected punishment	(1) US	(2) UK	(3) MA	(4) TR	(5) By cluster	(6) By country
Absolute negative deviation	0.304*** (0.106)	0.334*** (0.102)	−0.008 (0.165)	0.237*** (0.090)	0.325*** (0.072)	0.293*** (0.095)
Positive deviation	0.004 (0.083)	0.078 (0.083)	−0.352 (0.261)	−0.206** (0.104)	0.047 (0.063)	0.007 (0.081)
Absolute negative deviation × Non-WEIRD cluster					−0.171* (0.095)	
Positive deviation × Non-WEIRD cluster					−0.276** (0.113)	
Non-WEIRD cluster					0.048 (0.688)	
Absolute negative deviation × UK						0.069 (0.133)
Positive deviation × UK						0.087 (0.125)
UK						−1.093 (0.958)
Absolute negative deviation × MA						−0.288* (0.147)
Positive deviation × MA						−0.250 (0.177)
MA						−0.040 (1.000)
Absolute negative deviation × TR						−0.054 (0.125)
Positive deviation × TR						−0.233* (0.131)
TR						−0.608 (0.921)
Socio-economic control variables	No	No	No	No	Yes	Yes
Constant	−4.220*** (0.980)	−4.441*** (1.086)	−5.554** (2.465)	−3.850*** (1.050)	−9.832*** (2.587)	−9.358*** (2.652)
Pseudo R^2	0.05	0.09	0.02	0.09	0.7	0.08
No. obs. (Clusters)	318 (106)	264 (88)	240 (80)	258 (86)	1080 (360)	1080 (360)

Notes: Tobit coefficients. Robust *SE* clustered on individuals. Socio-economic control variables: age, female, urban background, middle class, single child, economics/business student. In Col. 5-6, age is positive and significant. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

H. Accuracy of expected punishment

We investigate whether there are differences in the accuracy of expected punishment (defined as the deviation of expected punishment from actual punishment) across the four samples (Figure H.1). In the US sample, 48% of beliefs about punishment are correct and 40% over-predict punishment. In the UK sample, 47% of instances are correctly predicted and 44% are overestimated. In the MA sample, 70% of punishment actions are correctly predicted. Here, participants overestimate the number of punishment points in 22% of instances. In the TR sample, 51% of punishment actions are correctly predicted and 38% are overestimated. Comparing the accuracy of beliefs about punishment across subject pools reveals highly significant differences (Kruskal-Wallis test, $\chi^2(3) = 18.93$, $p < 0.001$). These differences stem from highly significant across-cluster differences (pooled Mann-Whitney $Z = -2.91$, $p = 0.004$) and highly significant differences within the non-WEIRD cluster (Mann-Whitney $Z = -2.91$, $p = 0.003$). We find a similar accuracy of beliefs within the WEIRD cluster (Mann-Whitney $Z = -1.25$, $p = 0.212$). These results hint at differences in the salience of punishment norms across countries. In the MA sample, punishment is less prevalent, and a large share of people correctly predict the number of punishment points they actually receive. Therefore, the norm to punish defectors seems to be relatively weak. In the UK sample, punishment of defectors is more severe, but people overestimate the use of punishment.

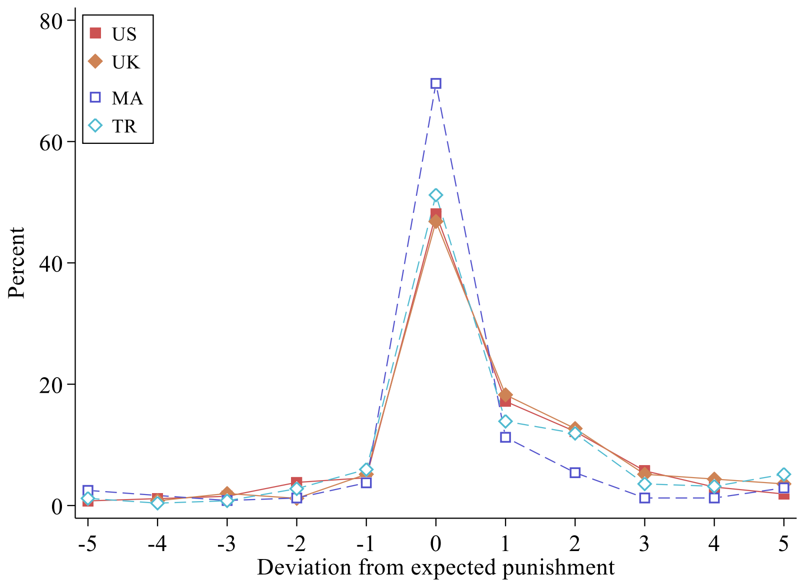


FIGURE H.1. Accuracy of expected punishment. If the deviation of expected punishment from the actual punishment equals zero, then beliefs are correctly predicted. A positive deviation shows an overestimation of punishment and a negative deviation indicates an underestimation.

I. Assessing the relative importance of beliefs and preferences for cooperation

To separate the relative effects of beliefs and preferences on contributions, we conduct a simulation inspired by Fosgaard, Hansen, and Wengstrom (2014). This procedure consists of the following steps:

1. We estimate the following OLS regression on the full sample:

$$Y_i = \beta_0 + \beta_1 PredC_i + \beta_2 b_i + \beta_3 Non\text{-}WEIRD_i + \epsilon_i.$$

In this regression equation the dependent variable Y_i is either the unconditional contribution in the D-Game or contribution in the P-Game. The independent variables are the predicted contribution $PredC_i$ based on the individual cooperative disposition a_i (contribution tables from the D-Game serve as a proxy for the cooperative disposition) and i's belief b_i about others' average contribution. The linear regression model reflects the assumption of a linear relationship between the independent variables and the dependent variable. For simplicity, we do not use a censored model as the predicted values very rarely fall outside the assumed contribution range. Table I.1 shows the regression results which are used to calculate the predicted contributions in the next step.

TABLE I.1. Explaining contribution behaviour.

Dependent variable: contribution	(1) D-Game	(2) P-Game
Predicted contribution	0.558*** (0.073)	0.318*** (0.059)
Belief	0.195** (0.078)	0.573*** (0.073)
Non-WEIRD cluster	−0.905* (0.487)	−2.042*** (0.498)
Constant	2.129*** (0.481)	1.614*** (0.523)
R^2	0.48	0.56
No. obs.	360	360

Notes: OLS estimation with robust *SE*. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

2. We randomly draw (with replacement) 90 beliefs from the respective game and 90 contribution tables (from the D-Game) from each cluster. 90 corresponds to the average country sample size in the experiment.

3. The following four sets of predicted contributions are calculated using the coefficients from Table I.1. Thus, this step will generate 90 predicted values for each of the four estimations explained below.
    a. *Estimation 1*: Calculate predicted contributions based on beliefs and contribution tables from the WEIRD cluster with the dummy variable set to zero (reflecting the WEIRD cluster).
    b. *Estimation 2*: Calculate predicted contributions based on beliefs from the non-WEIRD cluster, but contribution tables from the WEIRD cluster and the dummy variable set to zero.
    c. *Estimation 3*: Calculate predicted contributions based on beliefs and contribution tables from the non-WEIRD cluster, but with the dummy variable set to zero.
    d. *Estimation 4*: Calculate predicted contributions based on beliefs and contribution tables from the non-WEIRD cluster and with the dummy variable set to one (reflecting the non-WEIRD cluster).
4. For each of the four predicted contributions from Estimations 1-4, we calculate the average over the 90 predicted values.
5. In a standard bootstrapping approach, Steps 2-4 are repeated 100 times. Figure I.1 shows the actual average unconditional contributions in the WEIRD cluster as well as the average predicted contributions obtained in the bootstrapping procedure.

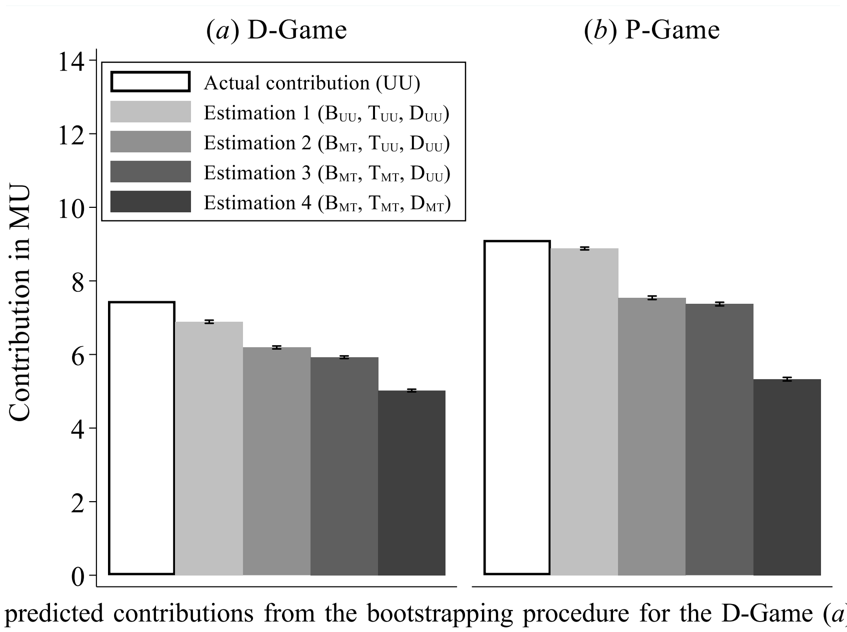


FIGURE I.1. Average predicted contributions from the bootstrapping procedure for the D-Game (*a*) and the P-Game (*b*). The error bars indicate ±1 bootstrapped *SEM*.

6. Finally, we decompose the total difference in predicted contributions between the two clusters in the belief effect, preference effect and remaining unexplained effect:
    a. *Estimation 4 – Estimation 1* = Total effect
    b. *Estimation 2 – Estimation 1* = Belief effect
    c. *Estimation 3 – Estimation 2* = Preference effect
    d. *Estimation 4 – Estimation 3* = Remaining unexplained effect

J. Instructions for the D-Game

This version of the instructions was used in the UK and the US. The Arabic and Turkish translations are available on request.

You are now taking part in an economic experiment. Depending on the decisions made by you and other participants, you can earn a considerable amount of money. It is therefore very important that you read these instructions with care.

These instructions are solely for your private use. **It is prohibited to communicate with other participants during the experiment.** If you have any questions, please raise your hand. A member of the experiment team will come and answer them in private. If you violate this rule, you will be dismissed from the experiment and you will forfeit all payments.

During the experiment, we will not speak in terms of Pounds, but in Guilders. At the end your entire earnings will be calculated in Guilders. The total amount of Guilders you have earned will be converted to Pounds at the following rate:

1 Guilder = 0.20 Pounds

After this experimental session, your entire earnings from the experiment will be paid to you privately in cash.

At the end of the session, you will be asked to fill in a questionnaire. The answers you provide in this questionnaire are completely anonymous. They will not be revealed to anyone either during the experiment or after it. Furthermore, your responses to the questionnaires will not affect your earnings during the experiment.

The groups
At the beginning of the experiment, all participants will be randomly divided into groups of four. Apart from you, there will be three other members in your group. **You will not learn who the other people in your group are at any point.**

The decision situation
Each participant receives an endowment of **20 tokens**. You have to decide how many of these 20 tokens you will contribute to a group project, and how many you will keep for yourself. The three other members of your group have to make the same decision. They can also either contribute tokens to the project or keep tokens for themselves. You and the other members of the group can each choose any amount between 0 and 20 tokens to contribute (including 0 and 20).

The payoffs
The income of every member of the group is calculated in the same way. Your income consists of two components:

(1) The first component is the amount of tokens that you keep for yourself. Every token that you do not contribute to the project automatically belongs to you and earns you one Guilder.

(2) The second component is your personal return from the group project. For all of the tokens contributed to the project the following happens: the project's value will be multiplied by 1.6 and this amount will be divided equally among all four members of the group.

For example, if 1 token is contributed to the project, the project's value increases to 1.6 Guilders. This amount is divided equally among all four members of the group. Thus every group member receives 0.4 Guilders.

The following function illustrates your income in Guilders:

Your Total Income = 20 – *Your Contribution* + 0.4 × (*Group Project*)

In order to explain the income calculation we will give some examples. Please read them carefully. At the end of the introductory information, you will be asked to answer several computerised control questions which are designed to check that you have understood the decision situation.

Example 1
If each of the four members of the group contributes 0 tokens to the project, all four will receive an income from their private account of 20. Nobody receives anything from the project, because no one contributed anything. Therefore the total income of every member of the group is 20 Guilders.

Calculation of the total income of every participant: (20 - 0) + 0.4 × (0) = 20

Example 2
If each of the four members of the group contributes 20 tokens, there will be a total of 80 tokens contributed to the project. The income from the private account is 0 for everyone, but each member receives an income from the project of 0.4 × 80 = 32 Guilders.

Calculation of the total income of every participant: (20 - 20) + 0.4 × (80) = 32

Example 3
If you contribute 20 tokens, the second member 10 tokens, the third member 5 and the fourth 0 tokens, the following incomes are calculated:

Because the total contribution to the project is 35 tokens, everyone will receive 0.4 × 35 = 14 Guilders from the project.

You contributed all your 20 tokens to the project. You will therefore receive 14 Guilders in total at the end of the experiment.

The second member of the group also receives 14 Guilders from the project. In addition, she receives 10 Guilders from her private account, because she contributed 10 tokens to the project. Thus, her total income is 24 Guilders altogether.

The third group member receives 14 Guilders from the project as well. Additionally, this group member will receive 15 Guilders from her private account. The total income therefore adds up to 29 Guilders.

The fourth member of the group, who did not contribute anything, also receives the 14 Guilders from the project and additionally the 20 Guilders from the private account, which means her total income is 34 Guilders.

Calculation of your total income: (20 - 20) + 0.4 × (35) = 14
Calculation of the 2nd group member's total income: (20 - 10) + 0.4 × (35) = 24
Calculation of the 3rd group member's total income: (20 - 5) + 0.4 × (35) = 29
Calculation of the 4th group member's total income: (20 - 0) + 0.4 × (35) = 34

Example 4
The three other members of your group contribute 20 tokens each to the project. You do not contribute anything. In this case the incomes will be calculated as follows:

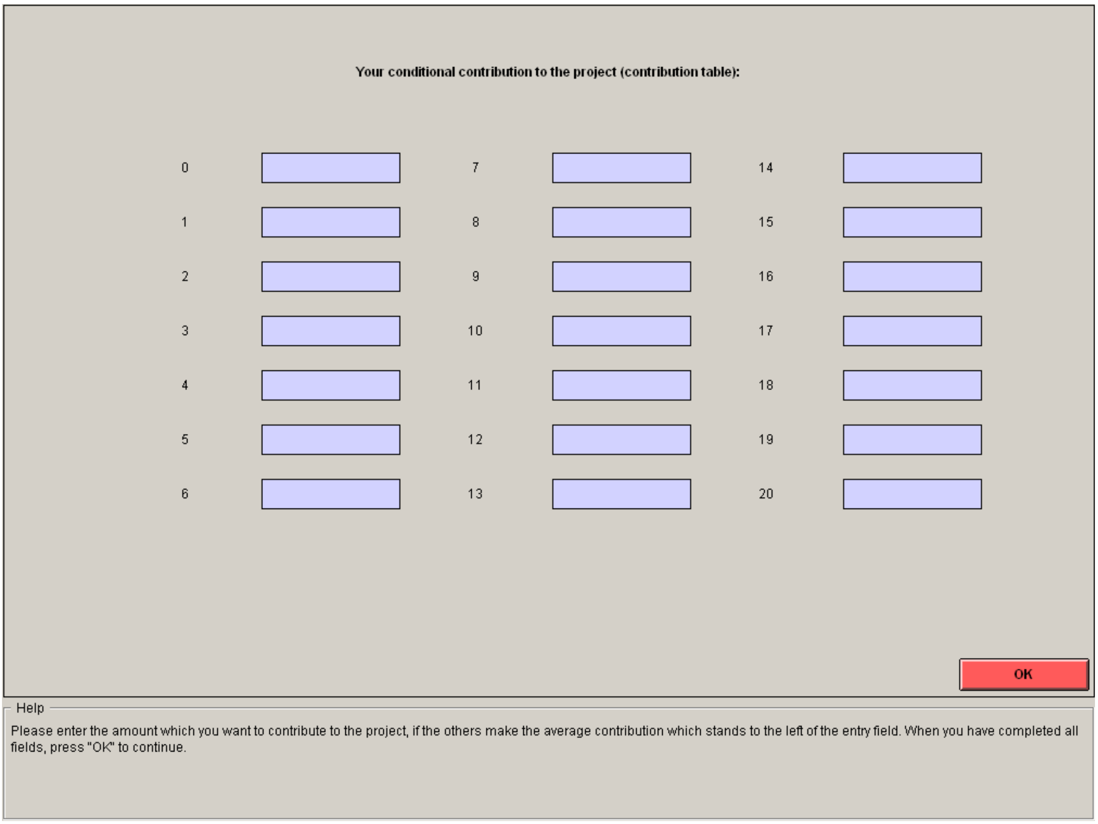
Calculation of your total income: (20 - 0) + 0.4 × (60) = 44
Calculation of the total income of each other group member: (20 - 20) + 0.4 × (60) = 24

The experiment
The experiment is based on the decision situation just described to you, conducted **only once**. In this experiment you will make two types of decisions: an **unconditional contribution** and filling in a **contribution table**.

When making your **unconditional contribution**, the following screen will appear:

As mentioned above, your endowment in the experiment is 20 tokens. You have to decide how many tokens you contribute to the project by typing a number between 0 and 20 (including 0 and 20) in the box. This box can be reached by clicking on it with the mouse. By deciding how many tokens to contribute to the project, you automatically decide how many tokens you keep for yourself. After entering the amount of tokens you want to contribute you must click on the "*OK*" button. Once you have done this, your decision can no longer be revised.

Your second task is to fill in a **contribution table** on the following screen:



The contribution table indicates **how many tokens you want to contribute to the project for each possible average contribution of the other group members** (rounded to the nearest integer). The table allows for conditioning your contribution on that of the other group members.

The numbers to the left of the input fields are the possible average contributions of the **other** group members (rounded to the nearest integer). You have to enter how many tokens you want to contribute to the project - conditional on the indicated average contribution of the other group members. **You must enter a number between 0 and 20 (including 0 and 20) into each box.**

For example, in the first box you enter the amount of tokens you want to contribute to the project in case the average contribution to the project of the other three group members is 0 tokens. In the next boxes you enter how much you contribute for an average contribution of 1, 2, 3, … tokens. After entering your decisions, you must click on the "*OK*" button.

After all participants of the experiment have made an unconditional contribution and have filled their contribution table, a random mechanism will select one member from every group. For **this group member, the contribution table** will be used to determine the contribution to the project. Whereas for **the other three group members, their unconditional contributions** will define the amount of tokens they add to the project.

You will not know whom the random mechanism will select before you make your unconditional contribution and fill in the contribution table. Therefore you must think carefully about both decisions. Either of them could determine your actual contribution to the project.

Example 5
Suppose that the **random mechanism selects you**; and that the other three group members made unconditional contributions of 0, 2, and 4 tokens, respectively. The average contribution of these three group members is, therefore, 2 tokens. If you indicated in your contribution table that you will contribute 1 token if the others contribute 2 tokens on average, then the total contribution to the project is given by 0 + 2 + 4 + 1 = 7 tokens. Each group member would, therefore, earn 0.4 × 7 = 2.8 Guilders from the project plus their respective income from their own private account. If, instead, you indicated in your contribution table that you would contribute 19 tokens if the others contribute 2 tokens on average, then the total contribution of the group to the project would be given by0 + 2 + 4 + 19 = 25 tokens. Each group member would earn 0.4 × 25 = 10 Guilders from the project plus their respective income from their own private account.

Example 6
Suppose that the **random mechanism does not select you**; and that your unconditional contribution is 16 tokens, while those of the other two group members not selected by the random mechanism are 18 and 20 tokens respectively. Your average unconditional contribution and that of these two other group members is, therefore, 18 tokens. If the group member whom the random mechanism did select indicates in her contribution table that she will contribute 1 token if the other three group members contribute on average 18 tokens, then the total contribution of the group to the project is given by 16 + 18 + 20 + 1 = 55 tokens. Each group member will therefore earn 0.4 × 55 = 22 Guilders from the project plus their respective income from their own private account. If, instead, the randomly selected group member indicates in her contribution table that she contributes 19 if the others contribute on average 18 tokens, then the total contribution of the group to the project is 16 + 18 + 20 + 19 = 73 tokens. Each group member would therefore earn 0.4 × 73 = 29.2 Guilders from the project plus their respective income from their own private account.

The random mechanism
Each group member is assigned a Group Member ID between 1 and 4, which denotes this participant's number inside her group. Moreover, participant number 2 was randomly selected at the very beginning of the experiment. This participant will draw a ball from an urn after all participants have made their unconditional contribution and have filled out their contribution table. Each ball in the urn has a different colour and each colour corresponds to a Group Member ID: orange = 1, blue = 2, yellow = 3, green = 4. The resulting number will be entered into the computer. If your Group Member ID is drawn, then your contribution table will determine your contribution to the project. For all other members of your group, the unconditional contributions will be relevant. Otherwise, your unconditional contribution determines your contribution.

If you have any questions, please raise your hand and a member of the experiment team will come and answer them in private.

K. Instructions for the P-Game

This version of the instructions was used in the UK and the US. The Arabic and Turkish translations are available on request.

You are now taking part in a second experiment. Your payoff from this experiment is completely unrelated to the decisions you have made in the previous one. The money you earn in this experiment will be added to what you earned in the first experiment. As before the Guilders you have earned will be converted to Pounds at the following rate:

1 Guilder = 0.20 Pounds

As in the previous experiment, all participants will be randomly divided into groups of four. However, the composition of the group is entirely new. **You will not learn who the other people in your group are at any point.**

The decision situation

The decision situation is the same as the one described on the first instruction sheet: Each participant receives an endowment of **20 tokens**. You have to decide how many of these 20 tokens you contribute to a group project and how many you keep for yourself. The three other members of your group have to make the same decision. However, this time you will make only an unconditional contribution to the project. There will be no contribution table.

After the contribution decision, there will be a **second stage**. At this stage, you will see how many tokens each of the other three group members has contributed to the project and their corresponding income from this contribution decision. Nonetheless, the identities of your group members will not be revealed at any stage. You can either **decrease** or **leave unchanged** the income of each other group member by assigning **deduction points** to them. The other group members can also decrease your income, by allocating deduction points to you, if they wish to do so.

Deduction points

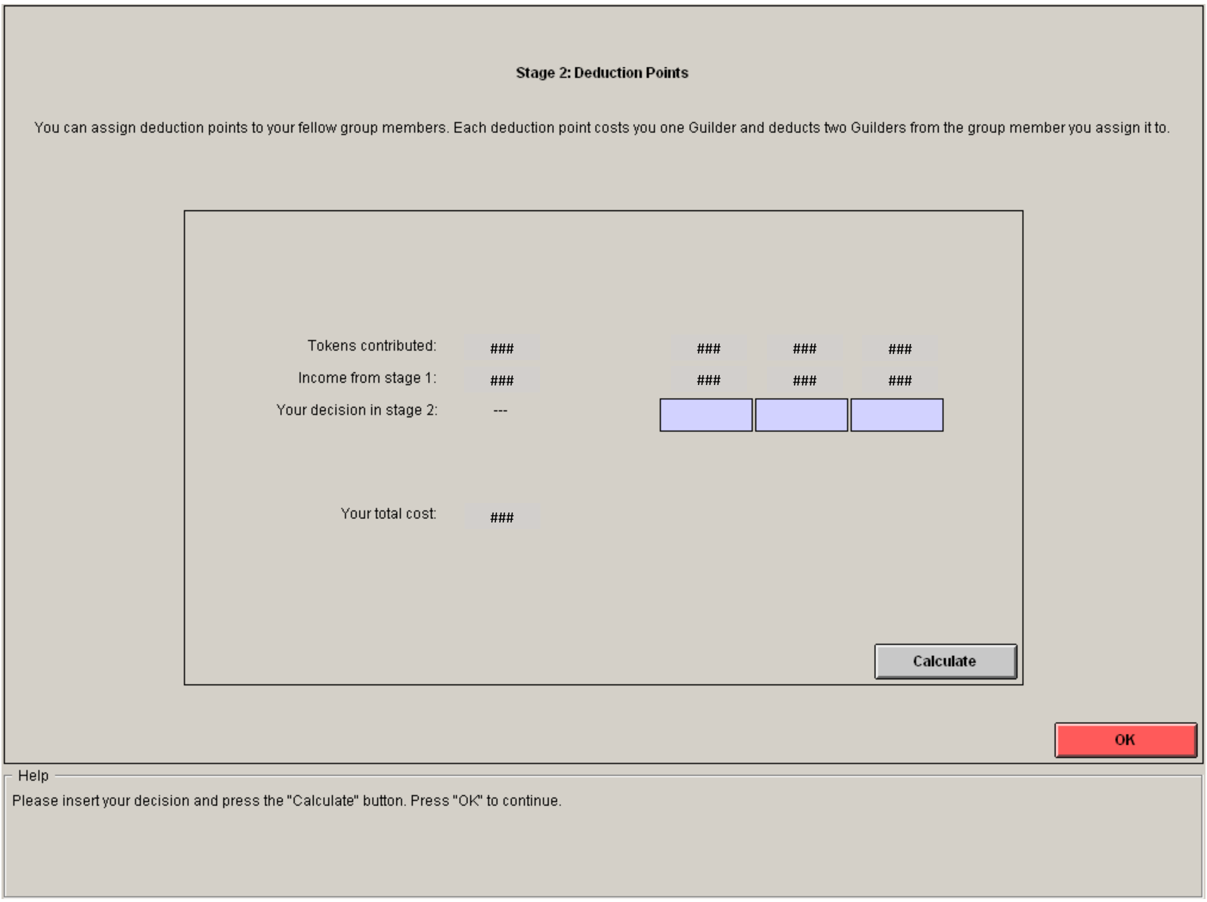
In stage 2, you can assign **between 0 and 5 deduction points to each other group member.** The maximum number of deduction points, you can allocate to the other group members together is therefore 15 deduction points.

For each deduction point that you assign, there is a cost to you of one Guilder. Thus, the total cost to you in Guilders of assigning deduction points to other group members is given by the total number of deduction points that you assign.

For each deduction point that you assign to a particular group member, you will decrease their income by 2 Guilders unless their income is already exhausted. For example, if you give a group member 2 deduction points, you will decrease this group member's income by 4 Guilders.

Your **own income will be reduced by 2 Guilders for each deduction point that is assigned to you** by the other three group members. If all of your income from the first stage of this experiment is exhausted, it cannot be reduced any further by other group members.

You will see the following screen at stage 2:



The column on the left shows your contribution and your income from the first stage. The other three columns indicate the contribution of your group members and their income from the first stage.

If you do not wish to change the income of the other group members, type "*0*" into the fields next to *"Your decision in stage 2"*. In case you want to assign deduction points, enter the number of deduction points you want to assign into this field. You must enter a decision into every field and press the "*Calculate"* button. This will display the cost of your decision. Until you press the "*OK*" button, you can still change your decision. To recalculate the costs after making a change, simply press the "*Calculate*" button again.

The payoffs

Your total income in Guilders from the two stages will be calculated as follows:

Your Income From Stage 1 = 20 – *Your Contribution* + 0.4 × (*Group Project*)

Total Income After Stage 2 = *Income From Stage 1* **(1)**
– 2 × (*Sum Of Deduction Points Assigned To You*) **(2)**
– (*Deduction Points Assigned By You*)
if (1) + (2) is greater or equal to 0.

Total Income After Stage 2 = 0 – (Deduction Points Assigned By You)
if (1) + (2) is less than 0.

Please note that your income in Guilders after stage 2 can be negative, if the cost of deduction points assigned by you exceeds your income from stage 1 less any reduction in your income caused by other group members.

However, at the end of the experiment and in addition to the calculation just given, you and the other members of your group will each receive a lump sum payment of **10 Guilders.** This payment is to cover losses that you could incur.

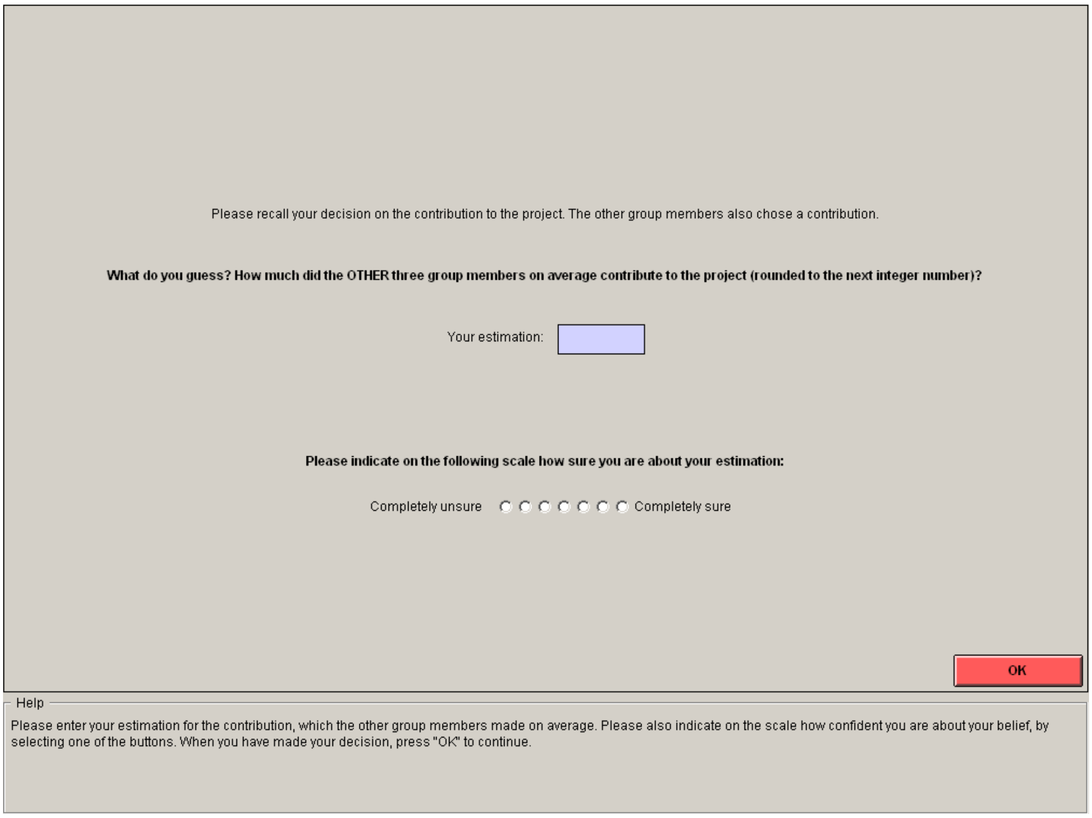
If you have any questions, please raise your hand and a member of the experiment team will come and answer them in private.

L. Decision screens not included in the instructions

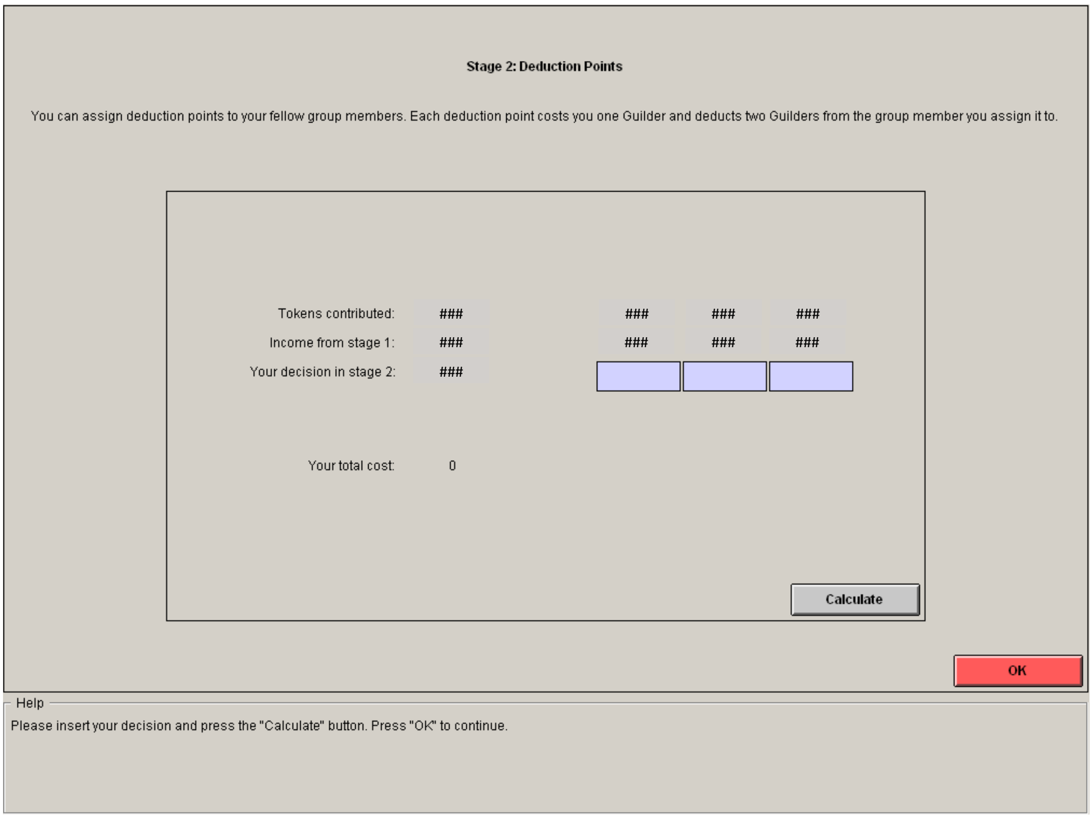
D-Game: Belief elicitation

P-Game: Contribution decision

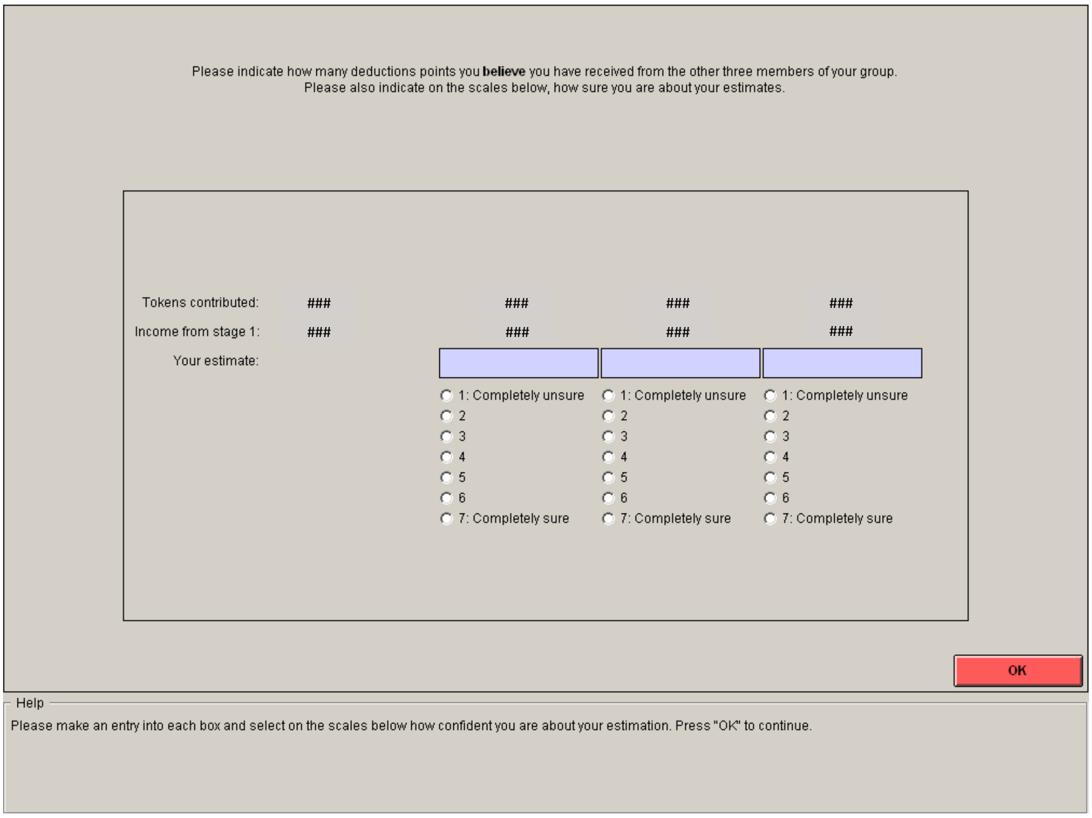
P-Game: Belief elicitation



P-Game: Punishment decision



P-Game: Elicitation of expected punishment



M. Socio-economic background questionnaire

We use the following questions from the socio-economic background questionnaire as controls in the regression analyses. The answer options of multiple-choice questions are provided in parentheses.

Your gender? (male/female)

How old are you?

What is your marital status? (single/married/widowed/divorced/prefer not to say)

If you are a student, what is the major area of your studies? (Natural Sciences/Engineering or Computer Science/Medical Science/Law/Humanities/Economics/Business Studies/Political Sciences/Social Sciences (other than Economics/Business)/Not a student). We combine the options “Economics” and “Business Studies” to create the dummy variable “Economics/ Business student”.

How large was the community where you spent the most time of your life? (up to 2,000 inhabitants/2,000 to 10,000 inhabitants/10,000 to 100,000 inhabitants/more than 100,000 inhabitants). We combine the options “10,000 to 100,000 inhabitants” and “more than 100,000 inhabitants” to create the dummy variable “Urban background”.

How many siblings do you have? We use the answers to this question to create the dummy variable “Single child”.

When you were 16 years of age, what was the income of your parents in comparison to other families in [*Country*]? (Far below average/Below average/Average/Above average/Far above average/Prefer not to say) We combine the options “Average”, “Above average” and “Far above average” to create the dummy variable “Middle class”.

We use the following question from the socio-economic background questionnaire to exclude participants who indicated that they were not citizens of the respective countries:

What is your nationality?

References

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2018). “Global evidence on economic preferences.” *The Quarterly Journal of Economics*, 133(4), 1645-1692.

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

Fischbacher, U., S. Gächter, and S. Quercia (2012). “The behavioral validity of the strategy method in public good experiments.” *Journal of Economic Psychology*, 33(4), 897-913.

Fosgaard, T. R., L. G. Hansen, and E. Wengstrom (2014). “Understanding the nature of cooperation variability.” *Journal of Public Economics*, 120(134-143.

Gächter, S. and J. F. Schulz (2016). “Intrinsic honesty and the prevalence of rule violations across societies.” *Nature*, 531(7595), 496-499.

Greif, A. (1994). “Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies.” *Journal of Political Economy*, 102(5), 912-950.

Herrmann, B., C. Thöni, and S. Gächter (2008). “Antisocial punishment across societies.” *Science*, 319(5868), 1362-1367.

Hofstede, G. H. and G. Hofstede (2001). *Culture’s consequences: Comparing values, behaviors, institutions and organizations across nations*. Thousand Oaks, CA: Sage Publications.

Inglehart, R., Haerpfer C., Moreno A., Welzel C., Kizilova K., Diez-Medrano J., Lagos M., Norris P., Ponarin E., and P. B. e. a. (eds.) (2014). “World Values Survey: Round Five - Country-Pooled Datafile Version: www.worldvaluessurvey.org/WVSDocumentation WV5.jsp.” Madrid: JD Systems Institute,

Inglehart, R. and C. Welzel (2005). *Modernization, cultural change, and democracy: The human development sequence*. Cambridge, UK: Cambridge University Press.

Kaufmann, D. and A. Kraay (2016). "Worldwide Governance Indicators." Dataset retrieved from: http://www.govindicators.org,

Kaufmann, D., A. Kraay, and M. Mastruzzi (2011). "The Worldwide Governance Indicators: Methodology and Analytical Issues." *Hague Journal on the Rule of Law*, 3(2), 220-246.

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

Molleman, L., F. Kolle, C. Starmer, and S. Gachter (2019). "People prefer coordinated punishment in cooperative interactions." *Nature Human Behaviour*, 3(11), 1145-1153.

Muthukrishna, M., A. V. Bell, J. Henrich, C. M. Curtin, A. Gedranovich, J. McInerney, and B. Thue (2020). "Beyond Western, Educated, Industrial, Rich, and Democratic (WEIRD) Psychology: Measuring and Mapping Scales of Cultural and Psychological Distance." *Psychological Science*, 31(6), 678-701.

Nikiforakis, N. and D. Engelmann (2011). "Altruistic punishment and the threat of feuds." *Journal of Economic Behavior & Organization*, 78(3), 319-332.

Weber, T. O., O. Weisel, and S. Gächter (2018). "Dispositional free riders do not free ride on punishment." *Nature Communications*, 9(2390).