Sanctions and international interaction improve cooperation to avert climate change

Gianluca Grimalda[I,II*], Alexis Belianin[III,IV], Heike Hennig-Schmidt[III,V], Till Requate[VI], Marina V. Ryzhkova[VII]

[I] Kiel Institute for the World Economy, Kiellinie 66, 24105 Kiel, Germany

[II] Centre for Global Cooperation Research at Duisburg-Essen University, Schifferstraße 44, 47059 Duisburg, Germany

[III] Higher School of Economics, Pokrovskiy Br, 11, 109028 Moscow, Russian Federation

[IV] IMEMO, Russian Academy of Sciences, 23, Profsoyuznaya Rd., 117997 Moscow, Russian Federation

[V] Bonn University, Regina-Pacis-Weg 3, 53113 Bonn Germany

[VI] Christian Albrechts University of Kiel, Wilhelm-Selig-Platz 1, 24118 Kiel (Schleswig-Holstein), Germany

[VII] Tomsk State University, 36 Lenin Ave, 63 4050 Tomsk City, Russian Federation

* Corresponding author: gianluca.grimalda@ifw-kiel.de

6 April 2022

This is a copy of the article published as:

Grimalda G, Belianin A, Hennig-Schmidt H, Requate T, Ryzhkova M V. "Sanctions and International Interactions Improve Cooperation to Avert Climate Change., *Proceedings Royal Society Biology.* **289** 20212174. https://doi.org/10.1098/rspb.2021.2174.

It is released under the "Green Open Access" policy made possible by the Royal Society.

Electronic supplementary material is available online at https://osf.io/6bcm7/.

Dataset and reproduction codes are available at: https://osf.io/r3a2x/

Abstract

Imposing sanctions on non-compliant parties to international agreements is advocated as a remedy for international cooperation failure. Nevertheless, sanctions are costly, and rational choice theory predicts their ineffectiveness in improving cooperation. We test sanctions effectiveness experimentally in international collective-risk social dilemmas simulating efforts to avoid catastrophic climate change. We involve individuals from countries where sanctions were shown to be effective (Germany) or ineffective (Russia) in increasing cooperation. Here we show that, while this result still holds nationally, international interaction backed by sanctions is beneficial. Cooperation by low cooperator groups increases relative to national cooperation and converges to the levels of high cooperators. This result holds regardless of revealing other group members' nationality, suggesting that participants' specific attitudes or stereotypes over the other country were irrelevant. Groups interacting under sanctions contribute more to catastrophe prevention than what would maximise expected group payoffs. This behaviour signals a strong propensity for protection against collective risks.

Background

Humanity is faced with a wide range of threats involving the possibility of catastrophic collective losses. Such threats require international cooperation across widely different cultures [1]. While cooperation may be sustained by direct and indirect reciprocity [2] in small or culturally cohesive groups [3], cooperation in large groups of unrelated individuals is typically parochial; that is, it favours others perceived as belonging to one's own group at the expense of others perceived as belonging to other groups [4-5]. Since nationality is one of the strongest sources of parochial attachment [6], international cooperation seems to be at risk [7]. Some scholars and policymakers have proposed to introduce substantial and credible sanctions – trade sanctions in particular – for countries that do not comply with international agreements as a way to improve cooperation [8-9]. Sanctions could take the form of increased tariffs on imported goods from countries not complying with international agreements. Yet, applying sanctions may trigger a second-order cooperation problem [9-10]. Sanctioning is generally costly to the party applying sanctions, thus each party will prefer to free ride on others' sanctions. Nevertheless, individuals, just like countries [1], seem to favour sanctioning policies, even when this is costly to them [11-13].

Climate change is perhaps the most severe existential threat facing humanity. Currently, international cooperation falls critically short of the levels necessary to mitigate climate change [14]**.** Sanctions have been proposed as a possible remedy to the current stalemate. The "climate club" proposal hinges upon the idea that countries not complying with a climate agreement suffer a penalty in the form of increased tariffs from countries belonging to the club [15]. This approach seems promising. Even if the Paris agreement is not a climate club as it does not allow for formal sanctioning, the number of climate provisions introduced in trade agreements is, in fact, increasing [16].

We designed an experiment to examine the effectiveness of sanctions for increasing international cooperation in interaction mimicking costs and incentives to prevent collective losses. Our experiment builds on the Collective Risk Social Dilemma [17] (CRSD; See Supplementary Materials (SM): Supplementary Notes: Section S7 for an abbreviation list). We modify the CRSD by introducing a sanctioning stage that reflects the characteristics of trade sanctions applied to international agreements. Controlled experimental evidence on sanctioning in inter-cultural contexts is rare [18] and lacking in international contexts. We involve participants from two countries - Germany and Russia – epitomising cultural areas where sanctions have been found to work or fail, respectively, as mechanisms to increase cooperation [19-20]. This approach puts the potential impact of sanctions as a method for underpinning international cooperation to a severe test.

Scientists classify climate change as having both a "gradual" and a "catastrophic" component. The former refers to incremental changes in underlying factors that continuously alter the climate, such as the progressive rise in sea levels. Catastrophic climate change refers to structural changes in ecosystems triggered by temperatures exceeding a "tipping point" and leading to irreversible change [21]. Examples are the collapse of the Amazon forest or the loss of ice sheets. The CRSD used in our experiments captures in a stylized way the potential gains and losses underlying catastrophic climate change [17]. Groups of individuals are faced with the possibility of losing part of their endowment if a random loss event occurs. To prevent such collective losses, individuals can contribute part of their monetary endowments to a collective fund that reduces the probability of the loss event occurring. Possible losses are large, thus simulating a major catastrophe in the offing. The consensus among scientists is that if temperatures increase less than 2˚C from pre-industrial levels, no catastrophic loss will

occur. We call this level the "certain safety threshold". On the other hand, an increase by more than 5˚C by 2100 – which would occur in a "business-as-usual" scenario – will certainly trigger catastrophic climate change [21]. We call this the "certain unsafety threshold". There is, however, uncertainty over which temperature level will actually trigger catastrophic climate change within the 2˚C–5˚C range [21-22]. We model uncertainty about the actual temperature level associated with this "catastrophe tipping point" using a uniform distribution over the interval bounded by the "certain safety" and the "certain unsafety" thresholds [22]. Collective loss is thus avoided with a probability proportional to the total amount of money that the group invests in the collective fund, relative to the amount of investment needed to achieve the certain safety threshold.

Experimental Methods

Participants were involved in the CRSD at either the National or the International level, with sanctions being possible (S-treatments) or not possible (NS-treatments). Groups of six participants interacted in the CRSD, three of whom were university students in one city and three in another. In National treatments, participants were informed that the other city was located in the same country. In International treatments, one city was in Germany and one in Russia. The International treatments were conducted under two different settings: In the Open (O-)treatments, German and Russian participants were informed that the other city was located either in Russia or in Germany. In the Blind (B-)treatments, participants were not made aware that participants from the other city were actually from another country [23]. Therefore, in both the National and the International Open treatments participants were made aware that the other city was located either in the same country or in another country, although the exact location of the other city was never disclosed. Participants were citizens of their country of residence. (See SM: section S1.1, S1.2 for participants' demographic and cultural characteristics).

Behaviour in international interaction may be affected by prejudice and stereotypes about foreigners [17], by national pride, or by the desire to outperform the other group [24]. Our experimental design permits the comparison of outcomes between the case where such prejudices or inter-group motivations may affect behaviour – i.e., in the Open treatments – and the case where prejudice or inter-group motivations are limited by construction – i.e., in the Blind treatments. Ex-post questionnaire data confirm that our treatment manipulation worked because a large majority of participants in the Blind treatments believed that they were interacting with participants from the same country (Table S4). The outcomes of the Blind treatments can thus be attributed solely to the effect of participants' choices, reducing the relevance of beliefs and motivations relative to interaction with foreigners. The eight experimental treatments are summarised in Table 1. The null hypotheses of equality of distribution across treatments of demographic and personal characteristics were not rejected in non-parametric tests within each country, except for Economics degree in Russia (Table S5). This entails that participants were in general not unevenly distributed across treatments with respect to such characteristics.

Treatment name	Within-Country/ International	Sanc-tions	Nationality revealed	N. of independent observations	N. of participants
GER_NAT_NS	National Germany	No	Yes	16	96
GER_NAT_S	National Germany	Yes	Yes	16	96
RUS_NAT_NS	National Russia	No	Yes	16	96
RUS_NAT_S	National Russia	Yes	Yes	16	96
INT_Blind_NS	International	No	No	16	96
INT_Blind_S	International	Yes	No	16	96
INT_Open_NS	International	No	Yes	16	96
INT_Open_S	International	Yes	Yes	16	96
Total				128	768

Note: GER_NAT=German participants; RUS_NAT=Russian participants; NS=No Sanctions; S=Sanctions; INT=International interaction.

Table 1 | Summary of Experimental Design

Participants interacted over ten periods with the same partners in real-time via the

Internet. Interactions were anonymous, but each group member could be identified by a number ranging from 1 to 6. Since participants knew that group members labelled from 1 to 3 were from one location while those labelled from 4 to 6 were from the other location, they could infer other group members' location from their numeric label. Each participant was endowed with 60 tokens in each period. Each token was worth €0.07 in Germany and 2.0 Ruble in Russia. Such levels ensured equivalent purchasing power across countries. In the NS-treatments, participants could contribute up to 50 tokens to a collective fund, the remaining 10 tokens being automatically added to their private accounts. If the sum of total contributions (C) to the collective fund matched or exceeded the certain safety threshold (T), there would be no loss to any player's private account. If, however, $C<T$ at the end of the ten periods, a loss of 75% to each player's private account would occur with probability $1-P$, where $P = min\left\{\frac{C}{T}; 1\right\}$ and P is the probability of loss avoidance (PLA) (SM: Fig. S4). P was the same for each group member. $C=0$ is the certain unsafety threshold. Individuals' private accounts at the end of ten periods would equal the total endowment of 600 tokens minus total individual contributions to the collective fund. Participants earned either the full amount in the private accounts at the end of the ten rounds if no loss occurred, or else a quarter of this amount.

The CRSD in the S-treatments took place in two stages. The first stage was identical to the NS-treatments. In the second stage, each group member could use the remaining 10 tokens from their endowments to reduce other group members' private accounts in each of the 10 periods. Tokens spent on such sanctions were deducted from the private account. This sanctioning system had a number of characteristics in common with typical sanctions in international trade agreements. First, sanctions were observable [25] as tariff systems are known to all relevant parties. Second, the number of tokens deducted from a sanctioned participant's account increased more than proportionally in the number of

tokens spent by other participants to sanction that participant. Similarly, the costs incurred by a country rise disproportionately as the number of sanctioning countries increases and as sanctions become more severe. The sanctioning cost structure is reported in Table S6. Final payoffs under the S-treatments were equal to those under the NS-treatments minus the sanctioning costs. At the end of each round of contributions, participants received information on each of the other group-members' contributions in all the previous rounds, as well as the current PLA determined by total contributions. In S-treatments, participants also received information on the sanctions assigned by a group member to any other group member.

We report details on the experiment protocol, measures adopted to ensure cross-country comparability, links to materials, and notes on determination of sample size, ethical approval, and generalisability of results in SM: section S4. Instructions and questionnaire are reported in SM: section 5-6.

Theoretical Benchmarks

We use two theoretical benchmarks to analyse this interaction. The Nash Equilibrium (*NE*) identifies the set of individual actions ensuring that each action is the best response to others' individual actions assuming that each agent maximises their own monetary payoff. By contrast, the cooperative solution (*CS*) takes the perspective of the entire group and maximises the total sum of expected monetary payoffs (SM: Section S1.3 for the derivation of the two solutions).

For low levels of T, both the *NE* and the *CS* prescribe the avoidance of losses with certainty (Fig. 1). For intermediate levels of T, individual and collective interests diverge as the *NE* prescribes progressively lower contributions, while the *CS* prescribes full loss avoidance. For T=2100, the threshold used in our experiment, the *NE* prescribes to contribute nothing – regardless of the individual's degree of risk aversion (Point A) –

while the *CS* for risk-neutral agents prescribes a PLA of 69% (Point B). If agents are risk-averse, the *CS* prescribes a higher PLA (Point C) than for risk-neutral agents, which is typically lower than certain loss avoidance (Point D). At *T*=2100, higher risk-aversion does not affect the *NE* (Point A).

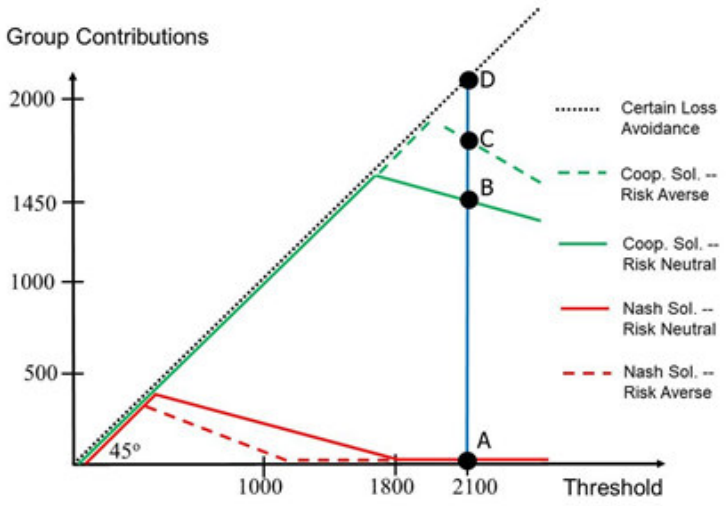


Fig. 1 | Nash Equilibria and Cooperative Solutions in the CRSD game for different levels of the certain safety threshold *T* (x-axis). The y-axis plots the group-level contribution for each solution.

It can be observed that higher risk aversion reduces individual contributions in the *NE* and increases those in the *CS* for internal solutions. These results are due to the fact that individuals and groups balance differently the effect of contributions on the *PLA*, on the one hand, and on the share of the private account that is earned even if the loss event occurs, on the other, differently. From an individual perspective, the effect of contributing more to the group account only slightly increases the *PLA*, while it for sure decreases the amount that is earned if the loss event occurs. Therefore, more risk-averse individuals will prefer to allocate more to the private account. From the group perspective, the effect of contribution on increasing the *PLA* is larger than in the

individual case, because the *CS* takes into account that a token contributed to the group increases the *PLA* for everyone in the group. Therefore, in the *CS* with more risk-averse individuals, more resources will be allocated to increasing the *PLA*.

Overall, the interaction implemented in our experiment had the typical characteristics of a social dilemma [26] with individual interests maximised by no contribution to the collective account and group interests maximised by positive contributions. Both the *CS* and the *NE* predict that no sanctioning should occur. This is so either because the *CS* already prescribes the collectively optimal levels of contribution, or because sanctioning of others is a second-order cooperation problem and rational self-interested individuals should not sanction (in the case of the *NE*).

Hypotheses and research questions

Our first two hypotheses concern the national treatments and the effectiveness of sanctions. Since the CRSD in our setting has not been investigated neither in an international setting nor under sanctioning conditions, we ground our hypotheses on other types of cross-national or international cooperation experiments. Most experimental studies on cooperation show that contributions start at an intermediate level between the *NE* and the *CS* and tend to get closer to the *NE*, without actually reaching it, as interactions go by. Two studies [6, 20] were concordant in finding that cooperation rates in Germany and Russia were no different from each other when sanctions were not available. Looking more generally at broad cultural areas, two studies found higher cooperation in countries from the "Protestant Europe" cultural group - to which Germany belongs [27] - than in the "Orthodox/ex-Communist" cultural group - to which Russia belongs [6, 19, 27-28] -, while another study found no difference between the two areas [20]. On the grounds of this evidence, we posit:

Hypothesis (HP)-1: Cooperation rates are not significantly different in Germany and Russia in National NS-treatments.

The same studies [19-20] found that sanctions were effective in increasing cooperation in the "Protestant Europe" cultural group - and in Germany in particular. By contrast, in the "Orthodox/ex-Communist" cultural group - and in Russia in particular - they were detrimental because of the widespread propensity to sanction high cooperators and to search for vengeance after having been sanctioned [19-20, 28-29]. We therefore posit:

HP-2: Sanctions are effective in increasing cooperation in Germany but not in Russia. Consequently, cooperation is overall higher in Germany than in Russia in National S-treatments.

As for international cooperation without sanctions, the parochial nature of human psychology [4-5] entails that cooperation should be lower in international than in national interaction, as national groups provide a strong source of attachment to individuals [30]. This is likely to be the case especially with repeated interaction because of the "bad apple" effect, i.e., the phenomenon whereby a few low cooperators in a group leads to a drastic reduction of willingness to cooperate with others [31]. Reduced cooperation in international interaction compared to national interaction has indeed been found [23,32]. However, other studies found no significant effect [33-34], with high-cooperators from one country possibly making up for the low cooperation rates by individuals from the other country [33]. Nevertheless, a large-scale study involving 42 countries found that ingroup bias between national and foreign groups was ubiquitous around the world [6]. On the grounds of this evidence, we posit:

HP-3: Cooperation is lower in International interaction than in National interaction in NS-treatments.

As for international interaction with sanctions, we already noted that sanctions have opposite effects on cooperation in Protestant European countries vis-à-vis Orthodox/ex-Communist countries [19-20,28-29]. It is therefore an open question whether sanctions will maintain the capacity to discipline low cooperators observed in Protestant European countries, whether they will trigger the retaliatory patterns observed in Orthodox/ex-Communist countries, or whether such two effects will cancel each other out. The overall effects on cooperation are also unclear. Given the lack of experimental evidence on sanctions in an international setting, we leave our research question open to two mutually exclusive hypotheses:

HP-4a: Sanctions are effective in increasing cooperation in International interaction above cooperation in International NS-treatments;

HP-4b: Sanctions fail to increase cooperation in International interaction above cooperation in International NS-treatments;

Results

Cooperation is substantial and averages the Cooperative Solution.

In contrast to the *NE* prediction, all groups achieved substantial levels of loss avoidance. We report the PLA achieved by each group in Fig. 2. Since the PLA is proportional to group members' total individual contributions, it is a measure of group-level cooperation. The PLA ranged from 12% to 100%, the grand mean being 70.1%, in line with the *CS* prediction for risk-neutral agents (Fig. 2). 86% of groups in the S-treatments achieved a PLA higher than the *CS*, with seven groups achieving full loss avoidance. Overall, the PLA was 18% lower in NS-treatments than S-treatments (d=-1.03; p<0.0001; N=128; d is Cohen's d; all tests are two-sided Wilcoxon-Mann-Whitney (WMW) tests unless otherwise indicated) and only 34% of groups in NS-treatments (as opposed to 86% in the S-treatments) exceeded the PLA prescribed by the *CS*. Neither contributions nor sanctions differed significantly between the two locations within each country (Tables S7-S8). Therefore, we consider aggregate national observations only. Moreover, the PLA achieved in the International Open treatments was very close in size - and not statistically significantly different - from the PLA in the International Blind treatments, particularly so in the NS-treatments (d=0.04; p=0.99; N=32), but also in the S-treatments (d=0.41; p=0.30; N=32). We report results for the International Open treatments (O-treatments) in the paper and those relative to the International Blind treatment in SM: Section: S1.4, unless results between the two treatments differ.

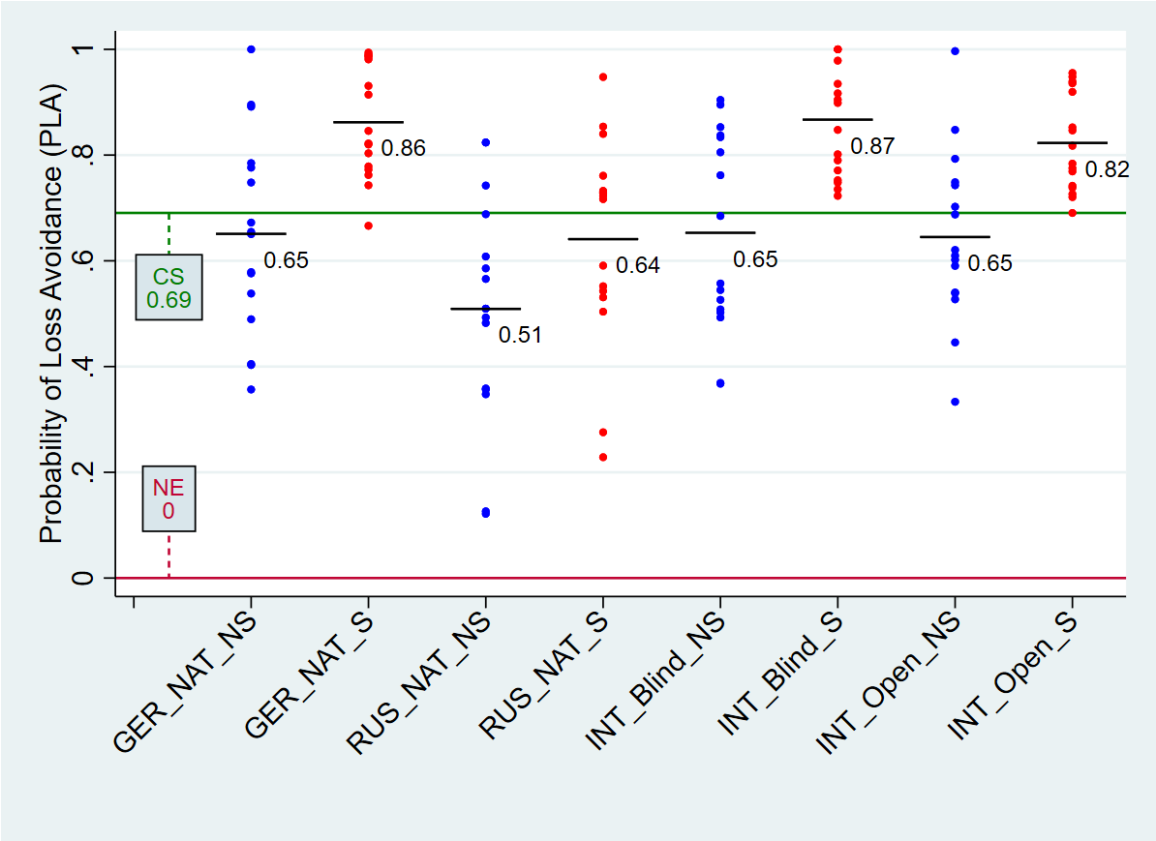


Fig. 2 | Probability of loss-avoidance for each group and treatment

Cooperation among Germans is higher than among Russians' in National treatments with sanctions.

We first assess Hypotheses 1-2. In National treatments without sanctions, German groups did not achieve significantly higher PLA than Russian groups at conventional levels, although the effect size was medium (d=0.71; p=0.08; N=32). Conversely, when sanctions were available, German groups did achieve significantly higher PLA than Russian groups in national interactions (d=1.39; p=0.0002; N=32). The PLA was significantly higher in the S-treatment than the NS-treatments in German national interactions (d= 1.37; p=0.0014; N=32). The increase in PLA in Russian national interaction in the S-treatment compared to the NS-treatment was not large enough to achieve statistical significance at conventional levels, although it had medium effect size

(d= 0.65; p=0.072; N=32). These results confirm Hypotheses 1-2 and are in line with previous comparative research [19-20,28-29].

Cooperation in International treatments without sanctions is not significantly different from cooperation in National treatments.

On the basis of Hypothesis 3, we would expect lower cooperation in International treatments without sanctions than in National treatments. This was however not the case. Without sanctions, mean PLA in the International O-treatment was 0.65, on a par with mean PLA in the German National treatment and higher than the mean PLA in the Russian National treatment (Fig. 2). Without sanctions, there was no statistically significant difference in PLA between the International O-treatment and either the German National NS-treatments (d= 0.03; p=0.99; N=32) or the Russian National NS-treatment, at conventional levels, (d=0.72; p= 0.055, N=32), although in the latter case the effect size was medium.

This result may be due to German participants having increased contributions in International NS-treatments to compensate for Russian participants' lower cooperation rates [33]. Alternatively, Russian participants may have increased their cooperation in International NS-treatments compared to National treatments. The latter alternative is supported by the data. German participants' cooperation levels were virtually the same as Russian participants' in the International O-treatment using two-tailed WMW matched-pairs sign rank tests (d=-0.18; signrank WMW: p=0.67; Fig. 3A).

We plot the evolution of average individual contributions per round broken down by nationality in SM: Fig. S5. It is noticeable that Russians' mean contributions in the International Open NS-treatment started off at a level virtually identical to what was

found in the Russian National treatment, but gradually caught up and matched up with Germans' mean contribution levels. However, a series of tests fail to reject the null hypothesis that Russians' contributions in the International Open NS-treatment come from the same distribution as in the National NS-treatment, except for periods 7 and 8 (Table S9 and SM: Section S1.4.1). Overall, we cannot reject, at conventional levels, the hypothesis that Russian participants contributed differently in the International NS-treatment and in the National NS-treatment, although the effect size is medium (d=0.57; p= 0.055, N=32). Moreover, there was no significant difference between Germans' contributions in the International Open NS-treatment and in the German National treatment neither in any interaction period (Table S9), nor across all periods (d=0.06; p= 0.70, N=32).

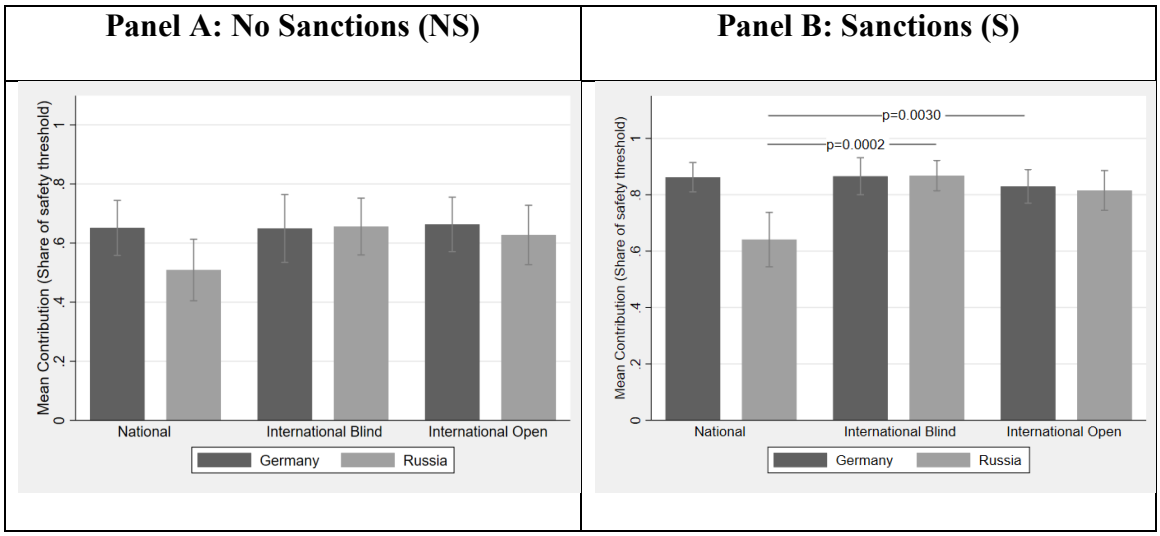


Fig. 3 | Average cooperation rates by nationality and treatment. Mean contributions to the collective fund as a share of the certain safety threshold. Error bars are 95% confidence intervals with bootstrapped standard errors (10,000 repetitions).

Cooperation in International treatments with sanctions is higher than without sanctions and, for Russians, higher than in National interaction.

Hypothesis 4 leaves open whether sanctions are effective in International treatments. We find that the PLA was significantly higher in the International Open S-treatment than in the Open NS-treatment (d= 1.35; p= 0.0012; N=32). Average PLA in International Open S-treatments was similar in size and not significantly different to that achieved in National German S-treatments (d=0.40; p= 0.22; N=32), but was significantly higher than average PLA in the Russian National S-treatment (d=1.18; p=0.003; N=32).

Again, this result may be due to Germans making up for Russians' lower cooperation in International treatments, or to Russians increasing their cooperation. As with the NS-treatment, we find that German participants' cooperation levels were virtually the same as Russian participants' in the International Open treatment (d=0.11; WMW signrank: p=0.86; N=16; Fig. 3, Panel B).

In the initial periods, contributions by Russian participants in International S-treatments started below German participants' contributions but quickly caught up as interactions continued (Fig. S5, Panels A-B). Non-parametric tests reveal that while Russian participants' contributions in the International Open S-treatment were not significantly different from Russian participants' contributions in the National S-treatment in periods 1 and 2, their contributions were significantly higher in International treatments than in National treatments in all subsequent periods (Table S9; SM: section S1.4.1). Conversely, the hypothesis of equality of distributions for contributions in International and National treatments was never rejected for German participants in any period. We can thus conclude that in International treatments, Russian participants' contributions quickly increased in comparison with the National treatment and converged

to German participants' contributions. Over the whole ten periods, contributions by Russians were significantly higher in the International O-treatment than in the National treatment with sanctions (d=1.00; p= 0.011, N=32). Conversely, there was no significant difference for Germans (d=-0.29; p= 0.45, N=32). These results support Hypothesis 4a and indicate that international cooperation with sanctions was overall beneficial because Russian participants achieved higher PLA, while PLA remained stable for German participants.

We decompose the treatment effects of cooperation by Russian and German participants in Table S10 and Fig. S6, pooling the two International treatments. Introducing sanctions in national interactions increases cooperation by 13% in comparison with the National NS-treatment. Remarkably, the same increase is effected without sanctions by "internationalising" interaction – i.e., having Russians interact with Germans. While neither of these effects is statistically significant, introducing sanctions in an international context increases Russian participants' cooperation by 20% in comparison to either the Russian National S-treatment or the International NS-treatment. As we noted, both these increases were statistically significant. We can thus conclude that, while sanctions alone and internationalisation alone brought about only marginal increases in cooperation by Russian participants, the combination of the two factors was necessary to significantly increase Russian participants' cooperation.

Little sanctioning suffices to spur cooperation.

Next, we analyse the mechanisms that made sanctions effective in increasing cooperation. Only 7% of the available endowment was spent on sanctions and in 70% of cases no sanction was administered (Fig. S7). Sanctions had a spike in the last period

when no counter-sanctioning was possible anymore (Fig. S10). This spike can only be accounted for as punishment for the previous or the present interactions, as it could not have any disciplinary function for the future.

Previous research found that sanctions are effective if people increase cooperation after having been sanctioned [20]. With an OLS econometric model controlling for period effects (see SM: section S1.5 for specification details), we compute the impact on the contribution made in the next period of having a token deducted through sanctioning in the current period. On average, a token deducted by sanctioning raised cooperation by 0.42 tokens in the next period ($p<0.001$; Table 2, Column 1), but the effect differed across treatments (Table 2, columns 2-5). Sanctioning was more effective in the German National treatment than in the Russian National treatment ($p=0.005$; Table 2, Column 3). Sanctioning in the International Blind treatment was as effective as sanctioning in the German National treatment ($p=0.76$, Table 2, column 4) and was significantly more effective than in the International O-treatment. Hence, sanctions lost part of their effectiveness when nationality was revealed to participants than when it was concealed from them.

DEPENDENT VARIABLE:	$Contribution_t - Contribution_{t-1}$				
	(1)	(2)	(3)	(4)	(5)
CONDITIONS	Estimated coefficients for $Sanction\_Loss_{t-1}$		Difference between RUS_NAT_S and other conditions' coefficients	Difference between GER_NAT_S and other conditions' coefficients	Difference between INT_B_S and INT_O_S
All conditions	0.42***				
	[0.10]				
	<0.001				
RUS_NAT_S		0.36***			
		[0.07]			
		<0.001			
GER_NAT_S		0.66***	-0.31***		
		[0.09]	[0.11]		
		<0.001	*0.005*		
INT_B_S		0.72***	-0.36*	-0.052	
		[0.14]	[0.16]	[0.17]	
		<0.001	*0.024*	*0.76*	
INT_O_S		0.29†	0.07	0.37*	0.43*
		[0.15]	[0.15]	[0.17]	[0.21]
		0.056	*0.69*	*0.032*	*0.041*
Observations	3,456	3,456	3,456	3,456	3,456
Number of participants	384	384	384	384	384
R^2_within	0.0887	0.105			
R^2_between	0.0436	0.0417			
R^2_overall	0.0748	0.0856			

Table 2 | Econometric analysis of the impact of sanctions on cooperation.

We fitted an OLS estimator to a model having as dependent variable the variation in Contribution between period t and t-1, for t=2,…10. The Table reports the estimated coefficients for tokens lost due to sanctions by other group members in the previous period ($Sanction\_Loss_{t-1}$) across all treatments (column 1) and for each treatment (column 2). Columns 3-5 report the results of t-tests over the null hypothesis that a certain treatment coefficient is different from the coefficient of another treatment. The full regression output and further specification details are reported in Table S11. *** p<0.001; ** p<0.01; * p<0.05; † p<0.10.

In order to understand the reasons why sanction effectiveness differed across treatments, we decompose sanctioning into *Prosocial* and *Antisocial* sanctioning (SM:

Section S1.4). We define Antisocial sanctions as instances in which an *ego* punished an *alter* who contributed no less than the group median, while Prosocial sanctioning is the residual category [25]. Antisocial sanctioning is puzzling because it targets individuals who are increasing social welfare in the group, but it has proved to be endemic in experiments with people from cultural areas classified as orthodox/post-communist [20,28-29].

We analyse treatment differences between the International O-treatment and the National treatments in the individual propensity to sanction through a Poisson regression controlling for period effects (see SM: S1.6 for the specification details). We control for the counterpart's contribution level, which, as expected, is a strongly significant predictor of sanctioning. The higher the contribution, the lower the probability of being sanctioned (b=-0.07, p<0.001, N=19,200; b is the coefficient estimated in the Poisson regression). It turns out that Germans spent significantly *more* on prosocial sanctioning than Russians in National treatments (b=0.32, p=0.016, N=7,210), and spent significantly *less* on antisocial sanctioning than Russians (b=-0.80, p=0.001, N=11,990). These results are at the bases of why sanctioning was effective in increasing cooperation in German National treatments but not in Russian National treatments. Germans significantly increased their prosocial sanctioning in the O-treatment compared with the National treatment (Wald test on difference in coefficients – Wald henceforth: b=0.34, p=0.018, N=7,210), and so did Russians (Wald: b=0.87, p<0.001, N=7,210). As a result, we found no significant difference in prosocial sanctioning between Russians and Germans in the International O-treatment (b=0.21, p=0.16, N=7,210). Interestingly, Germans increased their antisocial sanctioning in the International O-treatment compared to the National treatments, while Russians decreased it, although in both cases the differences were at the margins of statistical significance (Wald: b=0.53, p=0.062, N=11,990 for Germans; Wald: b=-0.47,

p=0.077, N=11,990 for Russians). Overall, there were no significant differences between Russians and Germans in the International O-treatment with respect to antisocial sanctioning (b=-0.20, p=0.49, N=11,990), nor in overall sanctioning (b=-0.12, p=0.43, N=19,200). Hence, Russians seem to have converged to the same patterns as Germans' sanctioning in the International O-treatment.

In the SM: S1.5, we show that these results are robust to demographic controls. We also analyse the effect of being sanctioned regardless of the amount of sanctioning (Table S12).

Additional results

We analyse demographic effects in SM: section S1.5-1.7 and show that payoffs were significantly higher in NS-treatments than S-treatments in SM: section S1.8.

Discussion

Many fear that as global-level cultural heterogeneity, complexity, and institutional limitations make international cooperation even more difficult than the local or national variety [26,35-36], international cooperation will be unable to steer clear of a tragedy of the "global commons" [37-38]. Our results offer a glimmer of hope, indicating that the combination of sanctions and the internationalisation of interaction brings about net positive effects. When sanctions were available, groups that are normally high cooperators in national interactions did not decrease cooperation internationally, while groups that are normally low cooperators nationally increased their cooperation internationally.

Our results indicate that individuals do not necessarily act parochially in social dilemmas where people cooperate to reduce collective risk. Theoretically, it may be argued that ingroup identity may be fostered by the common threat of losing part of the private account if the loss event occurs [39]. Such a shared fate may make the common goal of avoiding the loss particularly salient, thus prompting individuals to substitute collective goals for individual goals [40]. One may conjecture that a common ingroup identity is more easily created in a CRSD than in a standard public goods game [41]. The possibility of collective loss reduces the absolute differences in expected payoffs between high and low cooperators. Perhaps, this aspect of the interaction also makes it possible to cement a stronger group identity than in linear public goods interactions.

It has to be noted that, with some rare exceptions [23,33], the available evidence on international interaction is limited to one-shot interactions [6,7,42]. The dynamic setting may have created additional motivations for cooperation. One theoretical account hinges upon the idea of quicker belief update by low cooperators than high cooperators. According to this account, Russian participants' initial beliefs on others' cooperation would be rooted in the cooperation rates observed in local and national environments and would thus be set at a relatively low level. After observing higher-than-expected cooperation in the initial periods of interaction, though, Russian participants involved in international interactions would then be quick to revise their beliefs on their counterparts' cooperation upwards. Consistently with a motivational model of conditional reciprocity [19,43-44], adjusting beliefs upwards would then prompt Russian participants to be more cooperative in international than in national interactions. As for high cooperators, they may, contrary to the bad apple effect, behave and be perceived as role models [45], whose behaviour is imitated by low cooperators. Sanctions were necessary to achieve the result of significantly higher cooperation for low cooperators than in national interaction. A

novel result of this study is that German high cooperators were as capable of disciplining low cooperators in international interactions as in national ones. This result further qualifies the characteristics of strong reciprocators in cooperation interaction [4-12].

It is an open question whether our results are specific to the German/Russian combination or could be generalised to other countries or other contexts. The similarity of results in Blind and Open treatments shows that it was the actual content of participants' actions rather than motivations linked to the specific nationalities involved, which determined the beneficial effects of international cooperation. This suggests that our results are not driven by awareness of the counterpart's nationality and may be generalised to other countries from the same cultural groups. It has to be noted that the effectiveness of sanctions was also found in a study between different ethnicities within Bosnia [18]. Moreover, the conflictual history between Germany and Russia suggests that international cooperation observed between these two populations may be a lower bound of what is the same in other countries from different cultural areas. On the other hand, while cross-cultural analysis of cooperation patterns shows a remarkable consistency of results within cultural groups [28], the variance of behaviour is higher both when sanctions are available compared to when they are not [28] and in international as opposed to national contexts [23]. Germany and Russia are at intermediate levels of cultural distance [46] and it is therefore unclear what may happen when cultural distance increases. These considerations suggest caution on the possibility of straightforward generalisation.

Though our experiment reproduces in a stylized fashion various features of the consequences of climate change and trade sanctions on individual earnings, the problem of "scalability" is apparent in connection with the outcome of our experiment [47-48]. Nonetheless, at a more fundamental level, our experiment can be seen as revealing the

willingness of the general population to abide by an agreement once an agreement has been reached [49], which is a fundamental feature of any international agreement [1]. In fact, individuals who cooperated in the experiment were also marginally more likely to conduct environmentally sustainable behaviour in real life, such as buying environmentally-friendly goods, saving water, participating in ecological movements, and recycling (SM: Table S14 and SM: section 1.7).

Despite these limitations, our findings give rise to some policy recommendations. First, establishing international teams at several levels of government to seek solutions for collective risk social dilemmas seems a promising strategy. In spite of conspicuous cultural differences, our international groups were no less cooperative than national groups when a common threat loomed. Secondly, our results suggest that sanctions can be used in international interactions to discipline people who would otherwise not cooperate and that they can do this without risking a spiral of retaliation and counter-retaliation. This evidence supports the view that international sanctions can lead to significant and robust changes in standards of conduct and should be used more extensively in international agreements, particularly in climate agreements. A concrete policy recommendation would be to institutionalise sanctions, as is suggested through so-called climate clubs, by enforcing rules for members through internal sanctioning mechanisms, and implicitly sanctioning non-members through measures such as carbon border tax adjustments [50]. Our study has also shown a preference for remarkably high levels of collective loss-avoidance, at rates exceeding those that would maximise payoffs (see SM: section S1.8). Such preferences should be taken into account by policy-makers.

References

1. Barrett, S. 2007 *Why cooperate? The incentive to supply global public goods*. Oxford, UK: Oxford University Press.

2. Nowak M, Sigmund, K. 2005 Evolution of indirect reciprocity. *Nature* **437** (7063), 1291-1298.

3. Handley, C, Mathew, S. 2020 Human large-scale cooperation as a product of competition between cultural groups. *Nature Comms.* **11**, 1-9.

4. Choi, J, Bowles, S. 2007 The coevolution of parochial altruism and war. *Science* **318***,* 636-640.

5. Bernhard, H, Fischbacher, U, Fehr, E. 2006 Parochial altruism in humans. *Nature* **442** (7105), 912-915.

6. Romano, A, Sutter, M, Liu, JH, Yamagishi, T, Balliet, D. 2021 National parochialism is ubiquitous across 42 nations around the world. *Nature Comms.* **12**, 1-8.

7. Dorrough, AR, Glöckner, A. 2016 Multinational investigation of cross-societal cooperation. *P. Natl. Acad. Sci. U.S.A.* **113**, 10836-10841.

8. Martin, LL. 1994 *Coercive cooperation: Explaining multilateral economic sanctions*. Princeton USA: Princeton University Press.

9. Keohane, R, Victor, D. 2016 Cooperation and discord in global climate policy. *Nat. Clim. Change* **6**, 570–575.

10. Przepiorka, W, Diekmann, A. 2013 Individual heterogeneity and costly punishment: a volunteer's dilemma. *Proc. Royal Soc. B* **280** (1759), 20130247.

11. Henrich, J, McElreath, R, Barr, A, Ensminger, J, Barrett, C, Bolyanatz, A,... Lesorogol, C. 2006 Costly punishment across human societies. *Science* **312**, 1767-1770.

12. Boyd, R, Gintis, H, Bowles, S, Richerson, PJ. 2003 The evolution of altruistic punishment. *P. Natl. Acad. Sci. U.S.A.* **100**, 3531–3535.

13. Gürerk, Ö, Irlenbusch, B, Rockenbach, B. 2006 The Competitive Advantage of Sanctioning Institutions. *Science* **312**, 5770, 108-111. (doi:10.1126/science.1123633)

14. IPCC. 2021 *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [Masson-Delmotte, V, Zhai, P, Pirani, A, Connors, S L, Péan, C, Berger, S, Caud, N, Chen, Y, Goldfarb, L, Gomis, MI, Huang, M, Leitzell, K, Lonnoy, E, Matthews, JBR, Maycock, TK, Waterfield, T, Yelekçi, O, Yu, R, Zhou, B (eds.)]. Cambridge, UK: Cambridge University Press.

15. Nordhaus, W. 2015 Climate clubs: Overcoming free-riding in international climate policy. *Am. Econ. Rev.* **105**, 1339-70.

16. Berger, A, Brandi, C, Bruhn, D 2017 *Environmental provisions in trade agreements: promises at the trade and environment interface.* Bonn, Germany: German Development Institute, Briefing Paper, **16**.

17. Milinski, M, Sommerfeld, R, Krambeck, HJ, Reed, F, Marotzke, J. 2008 The collective-risk social dilemma and the prevention of simulated dangerous climate change. *P. Natl. Acad. Sci. U.S.A.* **105**, 2291-2294.

18. Alexander, M, Christia, F. 2011 Context modularity of human altruism. *Science* **334**, 1392-1394.

19. Gächter, S, Herrmann, B. 2009 Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment. *Phil. Trans. R. Soc. B* **364**, 791–806.

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

21. Kriegler, E, Hall, JW, Held, H, Dawson, R, Schellnhuber HJ. 2009 Imprecise probability assessment of tipping points in the climate system. *P. Natl. Acad. Sci. U.S.A.* **106**, 5041-5046.

22. Barrett, S, Dannenberg, A. 2012 Climate negotiations under scientific uncertainty. *P. Natl. Acad. Sci. U.S.A.* **109**, 17372-17376.

23. Finocchiaro Castro, M. 2008 Where are you from? Cultural differences in public good experiments. *J. Socio-Econ.* **37**, 2319-2329.

24. Böhm, R, Rockenbach, B. 2013 The inter-group comparison–intra-group cooperation hypothesis: comparisons between groups increase efficiency in public goods provision. *PLoS ONE* **8**, e56152.

25. Nikiforakis, N. 2008 Punishment and counter-punishment in public good games: can we really govern ourselves? *J. Publ. Econ.* **92**, 91–112.

26. Ostrom, E, Burger, J, Field, CB, Norgaard, RB, Policansky, D. 1999 Revisiting the commons: local lessons, global challenges. *Science* **284**, 278-282.

27. Inglehart, R, Welzel, C. 2005 *Modernization, Cultural Change and Democracy* New York, USA: Cambridge University Press.

28. Gächter, S, Herrmann, B, Thöni, C. 2010 Culture and cooperation. *Phil. Trans. R. Soc. B* **365**, 2651-2661.

29. Gächter, S, Herrmann, B. 2011 The limits of self-governance when cooperators get punished: Experimental evidence from urban and rural Russia. *Eur. Econ. Rev*. **55**, 193-210.

30. Balliet, D, Wu, J, De Dreu, CK. 2014. Ingroup favoritism in cooperation: a meta-analysis. *Psychol*. *Bull*. **140**,1556–1581.

31. Gino, F, Ayal, S, Ariely, D. 2009 Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. *Psychol. Sci.* **20**, 393-398.

32. Fiedler, S, Hellmann, DM, Dorrough, AR, Glöckner, A. 2018 Cross-national in-group favoritism in prosocial behavior: Evidence from Latin and North America. *Judgm. Decis. Mak.* **13**, 42–60.

33. Carpenter, J, Cardenas, JC. 2011 An intercultural examination of cooperation in the commons. *J. Conflict Resolut.* **55**, 632-651.

34. Goerg, SJ, Hennig-Schmidt, H, Walkowitz, G, Winter, E. 2016 In wrong anticipation. Miscalibrated beliefs between Germans, Israelis, and Palestinians. *PLoS ONE* **11**, e0156998.

35. Schmidt, KM, Ockenfels, A. 2021 Focusing climate negotiations on a uniform common commitment can promote cooperation. *P. Natl. Acad. Sci. U.S.A.* **118,** e2013070118. (doi:10.1073/pnas.2013070118)

36. Tavoni, A, Dannenberg, A, Kallis, G, Löschel, A. 2011 Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *P. Natl. Acad. Sci. U.S.A.* **108**, 11825-11829.

37. Hardin, G. 1968 The tragedy of the commons. *Science* **162**, 1243-1248.

38. Jacquet, J, Hagel, K, Hauert, C, Marotzke, J, Röhl, T, Milinski, M. 2013 Intra-and intergenerational discounting in the climate game. *Nat. Clim. Change* **3**, 1025-1028.

39. Froehlich, L, Dorrough, AR, Glöckner, A, Stürmer, S. 2021 Similarity Predicts Cross-National Social Preferences. *Soc. Psychol. Personal. Sci.* **12**, 1486-1498.

40. DeCremer, D, Van Vugt, M 1999 Social identification effects in social dilemmas: A transformation of motives. *Europ. J. Soc. Psychol.* **29**, 871-893.

41. Ispano, A, Schwardmann, P. 2017 Cooperating over losses and competing over gains: A social dilemma experiment, *Game. Econ. Behav.* **105**, 329-348.

42. Dorrough AR, Froehlich L, Eriksson K. 2021 Cooperation in the cross-national context. *Curr Opin Psychol*. **44**, 281-285. (doi: 10.1016/j.copsyc.2021.10.006)

43. Fischbacher, U, Gächter, S, Fehr, E. 2001 Are people conditionally cooperative? Evidence from a public goods experiment. *Econ. Lett*. **71**, 397-404.

44. Hovi, J, Ward, H, Grundig, F. 2015 Hope or despair? Formal models of climate cooperation. *Environ. Res. Econ.* **62**, 665-88.

45. Gächter, S, Renner, E. 2018 Leaders as role models and 'belief managers' in social dilemmas. *J. Econ. Behav. Org*. **154**, 321- 334.

46. Muthukrishna, M, Bell, AV, Henrich, J, Curtin, CM, 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. *Psychol. Sci.* **31**, 678-701.

47. Al-Ubaydli, O, List, JA, Suskind, DL. 2017 What can we learn from experiments? Understanding the threats to the scalability of experimental results. *Am. Econ Rev.* **107**, 282-286.

48. Messner, D, Guarín, A, Haun, D. 2015 The behavioral dimensions of international cooperation. In *Global Cooperation and the Human Factor in International Relations* (eds. D. Messner, S. Weinlich), Abingdon, UK:Routledge) pp. 65-83.

49. Guala, F. 2012 Reciprocity: weak or strong? What punishment experiments do (and do not) demonstrate. *Behav. Brain Sci.* **35**, 1-15.

50. European Commission (2021): *Proposal for a Regulation of The European Parliament and of The Council establishing a carbon border adjustment mechanism*, Brussels, 14.7.2021. COM(2021) 564 final, 2021/0214 (COD), Document 52021PC0564. (https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52021PC0564; accessed 2022/01/18)

Acknowledgments

We thank Giulia Andrighetto, Francesco Bogliacino, Marilynn Brewer, Jaume García-Segarra, José-Alberto Guerra, Andrea Guido, Manfred Milinski, Rustam Romaniuc, Angelo Romano, Gari Walkowitz, and Joachim Weimann for helpful comments. We are especially grateful to all research assistants involved in this project, who are listed at SM: Section 8.

Authors' contribution

Conceptualization: all; Experimental data collection: all; Data curation: AB, HHS, TR, MR; Formal analysis: AB, TR; Funding acquisition: AB, GG, HHS, TR; Investigation: All; Project administration: All; Writing – original draft: GG; Writing – review & editing: All.

Funding

The work was supported by the Center for Global Cooperation Research at the University of Duisburg-Essen, Christian-Albrechts-University of Kiel, and by HSE within the framework of the Basic Research Program at the National Research University Higher School of Economics (HSE).

Data availability Statement:

The dataset generated and analysed during the current study and analyses codes are available in the project repository of the Open Science Foundation: https://osf.io/r3a2x/ and at Dryad: https://doi.org/10.5061/dryad.qv9s4mwgd

Sanctions and international interactions improve cooperation to prevent climate change

Gianluca Grimalda[I,II*], Alexis Belianin[III,IV], Heike Hennig-Schmidt[III,V], Till Requate[VI], Marina V. Ryzhkova[VII]

[I] Kiel Institute for the World Economy, Kiellinie 66, 24105 Kiel, Germany
[II] Centre for Global Cooperation Research at Duisburg-Essen University, Schifferstraße 44, 47059 Duisburg, Germany
[III] Higher School of Economics, Pokrovskiy Br, 11, 109028 Moscow, Russian Federation
[IV] IMEMO, Russian Academy of Sciences, 23, Profsoyuznaya Rd., 117997 Moscow, Russian Federation
[V] Bonn University, Regina-Pacis-Weg 3, 53113 Bonn Germany
[VI] Christian Albrechts University of Kiel, Christian-Albrechts-Platz 4, 24118 Kiel (Schleswig-Holstein), Germany
[VII] Tomsk State University, 36 Lenin Ave, 63 4050 Tomsk City, Russian Federation
* Corresponding author: gianluca.grimalda@ifw-kiel.de

Electronic Supplementary Materials

Table of contents

S1 Supplementary analysis

S1.1: Socio-economic background of participant pools

768 individuals participated in our study, half of them Germans and half Russians. The experimental sessions were conducted from November 2016 to February 2017 at the laboratories of the Higher School of Economics, Moscow, of the Tomsk State University of Control Systems and Radioelectronics (in the Russian Federation), and at Bonn University and Kiel University (in Germany).

Moscow is the capital of the Russian Federation and is located in Central Russia, its population being about 12.5 million inhabitants in the city area plus about 7.5 million in the Moscow region, which is an urbanised area near the capital city. Tomsk is the administrative center of Tomsk oblast (region) located in the southwest of Siberia and has about 580,000 inhabitants. Bonn was the capital of the Federal Republic of Germany from 1949 to 1990. It has about 330,000 inhabitants and is situated in the Federal State of North-Rhine Westphalia, located in West Germany. Kiel has a population of about 240,000 inhabitants and is the capital of the Federal State of Schleswig-Holstein, in Northern Germany.

Table S1 summarises participants' distribution across the four locations. About three quarters of German participants studied in the region where they were born (75% in both Kiel and Bonn). About 50% of Moscow participants are born in the Moscow area or Central Russia, while in Tomsk nearly 90% of participants are born in the regions east of the Urals. While the sample is evenly balanced across the two Russian locations, due to logistical constraints the Bonn laboratory was not available on some dates, hence we ran some extra sessions at the Kiel laboratory. Since we find no significant differences in behaviour between participants in the two locations within each country (see Table S7-S8), we do not believe that results are affected by the uneven distribution of observations between the two German locations.

Laboratory	Frequency	Percentage
Bonn	168	21.88
Kiel	216	28.12
Moscow	192	25.00
Tomsk	192	25.00
Total	768	100.00

Table S1: Distribution of participants across locations
Notes: This table reports the absolute and relative frequencies (%) of participants per participating laboratory.

An anonymous post-experimental questionnaire provides us with further socio-demographic details of our participant pools (see Table S2). We do not report income data due to a high percentage of missing data and implausible data entries.

German participants are older than Russian participants, reflecting the fact that enrolment at university typically occurs two years earlier in Russia than Germany. The gender distribution was balanced in the two countries: 49% of participants are females in Germany and 51% in Russia. Nearly all participants were not married. The language primarily spoken in families is German or Russian, respectively. Participation in academic exchange programs was comparable in both participant pools. In both countries around 40% of participants reported being Christians and about the same proportion reported being atheists. About one fifth of German participants studied Humanities and

Social Sciences, Mathematics and Natural Sciences, or Economics and Management each, while the majority of Russian participants majored in Management and Economics.

VARIABLES (average or %)	German	Russian
Age (years)	23.28 (3.77)	20.43 (2.92)
Female (%)	49.25	50.75
Height (cm)	175.27 (9.84)	172.93 (9.67)
Marital status: married (%)	1.05	3.65
National language spoken at home (%)	95.54	96.35
Academic exchange (%)	12.50	16.15
Religious denomination (%)		
Christians (Germany: Catholics, Protestants; Russia: Orthodox)	44.53	40.36
Atheists/Agnostics	43.49	44.53
Other	11.98	15.20
Participant's degree (%)		
Humanities and Social Sciences	24.74	17.45
Mathematics and Natural Sciences	22.66	19.01
Economics and Management	20.57	52.86
Other	32.03	10.68
Risk Tolerance	5.39 (2.07)	5.83 (2.13)

Table S2: Demographic characteristics of participants' pools
Note: The table reports the frequency observed for each characteristic in the German and Russian participant pool. Standard deviation for age, height, and risk tolerance is in parentheses. Other religious denomination includes e.g. Buddhist and Muslim in both countries, Orthodox Christians in Germany, and Protestant and Catholic in Russia. Other Participant's degree includes, e.g., medicine, law, psychology, theology. Risk Tolerance is measured through Q30 in Questionnaire (see ESM: section S6).

90% of German participants' fathers were born in the territory of the Federal Republic of Germany (FRG) or the former German Democratic Republic (GDR), 5% in the EU, and 5% in other European countries (including Russia), in Turkey, other Asian countries, Africa, Australia, or the US. The distribution is similar for participants' mothers (FRG or GDR: 91%; EU: 5%, other European countries incl. Russia, in Turkey,

Asia or Africa: 4%). Parents are from 29 different countries that partially overlap between fathers and mothers.

Nearly all fathers of Russian participants are born in the territory of the former USSR (99%), of whom at least 70% in the Russian Federation[1], and only 1% outside of the former USSR. The figures for mothers are nearly the same (former USSR: 99%; RF: at least 71%; outside of USSR: 1%). Parents are from 16 different countries with the parents' native countries mostly overlapping.

We classified participants' mothers' and fathers' highest level of education according to the following categories[1]: 1. Primary or lower education; 2. Secondary (lower or upper) education; 3. A-levels or post-secondary non-tertiary education; 4. Tertiary education (Bachelor, Master or Diploma degrees); 5. PhD or more than two diplomas or science degrees. Fig. S1 shows that the highest fraction of parents in the German participant pool has secondary education (39%), while the majority of Russian parents (56%) have an academic degree (Bachelor, Master or Diploma).

Russian parents' educational backgrounds regarding secondary and tertiary education are rather similar to the OECD data on educational attainment of 25 to 64 years old adults (21% vs. 24%, and 56% vs. 55%, respectively, see Fig. S1). This holds to some extent also for the German parents' academic education (32% vs. 27%). The fraction of German parents with secondary education is lower than the OECD data. In both countries, parents holding a PhD, or more than two diplomas or science degrees, are overrepresented compared to the OECD data.

[1] While 70% of Russians participants explicitly stated that their parents were born in Russia or the Russian Federation, 11.5% stated that their parents were born in the Soviet Union (USSR) without specifying whether their birthplace was located within the current boundaries of the Russian Federation or within one of the now independent states.

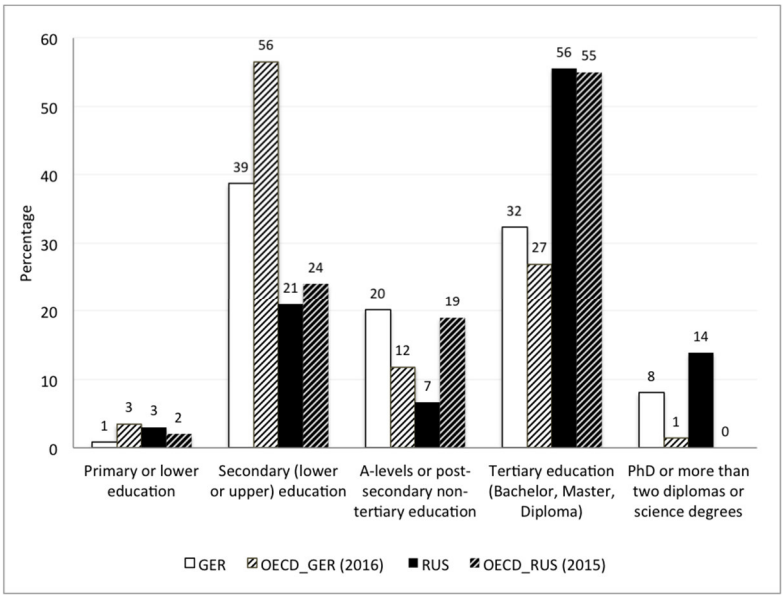


Fig. S1: **Parents' education and OECD data**[1] **on educational attainment of 25-64 year-old adults**. The figure reports the percentage distributions of German and Russian participants' mothers' and fathers' highest level of education (*N*=1,507). In addition, OECD data on educational attainment of 25-64 years old adults in 2016 are shown. Source: OECD data [51]: Indicator A1, Table A1.1: Educational attainment of 25-64 year-olds.

S1.2 Cultural variation in Russian and German populations

According to international surveys, Russian and German populations differ along many cultural traits. The Inglehart-Welzel world cultural map [27] ranks countries according to two scales (Fig. S2). The first scale contrasts Survival values - characterised by search for economic and physical security, a relatively ethnocentric outlook and low levels of trust and tolerance – and Self-expression values – in turn characterised by search for subjective well-being, self-expression, and quality of life. The second scale contrasts traditional values, which are centred around religion, deference to authority, traditional family values, and a nationalistic outlook, where secular-rational values have the opposite preferences to traditional values. Russia is a typical exponent of the "Orthodox Europe" group, scoring slightly below average in the Survival vs. Self-Expression Values scale and slightly above average in the Traditional vs. Secular Values scale. Conversely, Germany epitomises the "Protestant Europe" group, ranking among the top in both scales. The difference between the two countries appears particularly large on the Survival vs. Self-Expression Values scale rather than on the other dimension.

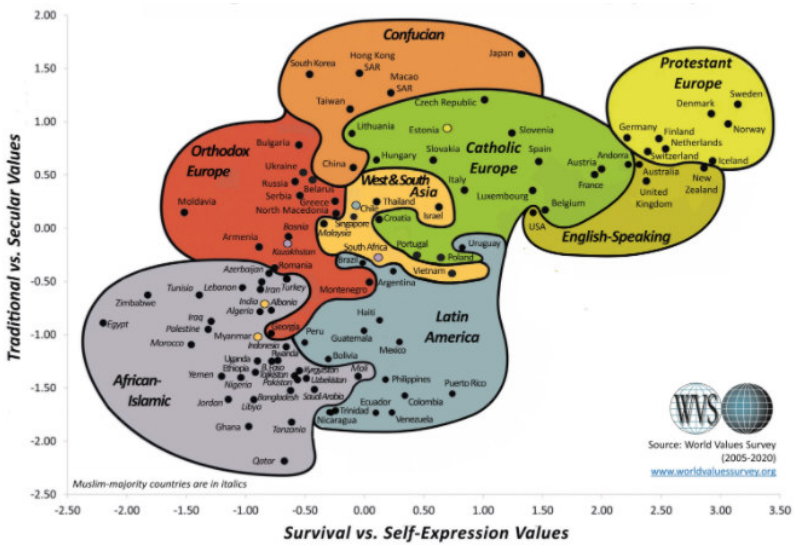


Fig. S2: Inglehart-Welzel cultural map.
Source: [52]

According to the Hofstede's six-dimension model [53] Russia ranks at the top positions on power distance, namely, "*the extent to which the less powerful members of institutions and organisations within a country expect and accept that power is distributed unequally*", while Germany is among the lower power distance countries (Fig. S3). If Germans attach high value to competition, achievement and success (labelled as "masculine" in the model), Russians score lower on this scale, as they attribute high value to caring for others and quality of life. While both countries score high in uncertainty avoidance, which spawns beliefs and institutions aiming to avoid uncertainty, Russia scores 30 points higher in this index. While German culture is classified as highly individualistic in Hofstede's model, Russian culture is ranked as collectivistic. The only dimension in which the countries are similar is long-term orientation, which is highly valued in both countries.

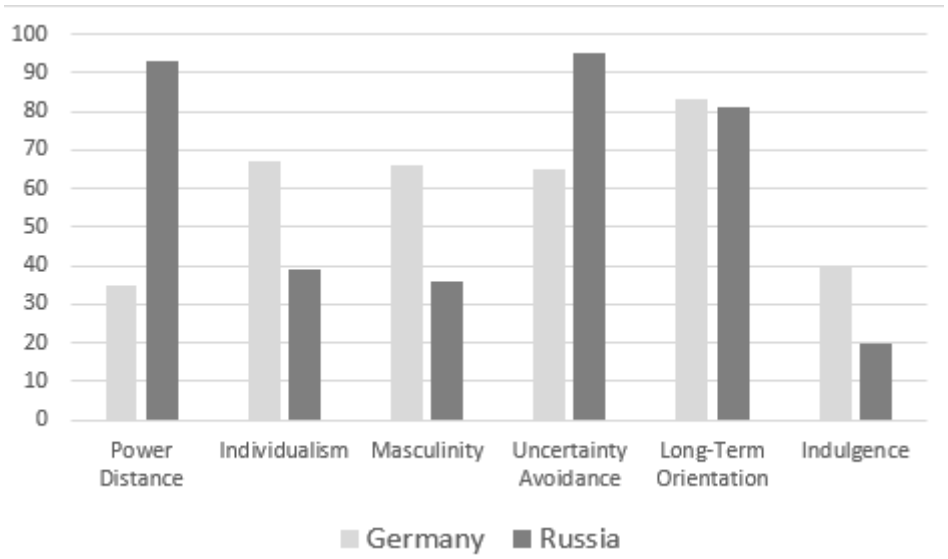


Fig. S3: Scores of Germany and Russia on Hofstede six-dimensions of national culture model.
Source: [54]

Finally, in a global survey of economic preferences [55], Russia and Germany appear relatively close in terms of risk taking, positive and negative reciprocity, altruism and trust, but very different on patience, Germans being more patient than Russians. Patience is also the category most strongly associated with economic prosperity.

Such differences in cultural traits are also reflected in our student sample. German and Russian participants held cultural beliefs on the acceptability of socially or morally relevant behaviours that were significantly different from each other for eight out of the ten dimensions being considered (Table S3). While participants from two German locations held significantly different beliefs only in one out of the ten dimensions considered, cultural differences were more extensive between Moscow and Tomsk, with seven significant differences out of ten. Finally, German participants were more worried that global warming represents a threat to them or their families than Russian participants.

1

City		Statis-tics	Benefits_Claim	Fare_Avoidance	Tax_Cheating	Bribe	Homosexuality	Prosti-tution
Bonn		Mean	1.71	2.50	1.54	1.66	4.79	3.29
		St. Dev.	0.89	1.02	0.76	0.92	0.71	1.15
Kiel		Mean	1.73	2.24	1.54	1.61	4.82	3.19
		St. Dev.	0.95	1.12	0.84	0.90	0.57	1.11
	Tests between German locations	Z-statistics	0.18	2.63**	0.55	0.66	-0.06	0.77
		p-value	*0.86*	*0.008*	*0.58*	*0.51*	*0.95*	*0.44*
Moscow		Mean	2.86	2.53	2.49	1.63	3.67	3.16
		St. Dev.	1.04	0.96	1.14	0.85	1.54	1.38
Tomsk		Mean	2.79	1.99	2.10	1.58	2.62	2.18
		St. Dev.	1.14	1.04	1.20	1.00	1.69	1.40
	Tests between Russian locations	Z-statistics	0.75	5.71***	3.694***	1.64	5.94***	6.784***
		p-value	*0.45*	*<0.0001*	*0.0002*	*0.1009*	*<0.0001*	*<0.0001*
	Tests between Germany and Russia	Z-statistics	-13.84***	1.22	-9.50***	0.71	15.37***	5.84***
		p-value	*<0.0001*	*0.22*	*<0.0001*	*0.4784*	*<0.0001*	*<0.0001*

2

City		Statistics	Divorce	Eutha-nasia	Suicide	Beat_ Wife	Global_Warming_Threat
Bonn		Mean	4.50	3.46	3.04	1.08	0.53
		St. Dev.	0.82	1.26	1.29	0.32	0.50
Kiel		Mean	4.61	3.51	3.11	1.09	0.56
		St. Dev.	0.69	1.09	1.27	0.41	0.50
	Tests between German locations	Z-statistics	-1.31	-0.04	-0.57	0.22	-0.56
		p-value	*0.19*	*0.97*	*0.57*	*0.83*	*0.58*
Moscow		Mean	4.30	4.08	2.67	1.27	0.32
		St. Dev.	1.08	1.18	1.53	0.56	0.47
Tomsk		Mean	3.43	3.18	1.89	1.28	0.50
		St. Dev.	1.33	1.49	1.30	0.72	0.50
	Tests between Russian locations	Z-statistics	6.77***	5.98***	5.35***	1.01	-3.63***
		p-value	*<0.0001*	*<0.0001*	*<0.0001*	*0.3108*	*0.0003*
	Tests between Germany and Russia	Z-statistics	7.80***	-2.53*	8.15***	-5.18***	3.72***
		p-value	*<0.0001*	*0.0115*	*<0.0001*	*<0.0001*	*0.0002*

Table S3: Differences in cultural traits between locations. The table reports mean and standard deviation of answers to questions tapping into cultural traits, taken from the World Value Survey. The text of the questions is reported in Section S6: Question 23. Answers were given on a 1-5 scale where 1 means "Never justifiable" and 5 means "Always justifiable". The questions inquired about a participant's acceptance of claiming government benefits to which one is not entitled (Benefits_Claim), avoiding a fare on public transport (Fare_Avoidance), cheating on taxes if one has the chance (Tax_Cheating), someone accepting a bribe in the course of their duties (Bribe), homosexuality (Homosexuality), prostitution (Prostitution), abortion (Abortion), divorce (Divorce), euthanasia (Euthanasia), suicide (Suicide), and of a man beating his wife (Beat_wife). We

also report means and standard deviation to Question 17, asking participants to state whether they think that global warming will pose a serious threat to them or their family in their lifetime (Global_Warming_Threat). Answers to the last question were dichotomous (yes/no). We also report z-statistics and p-values of two-tailed Wilcoxon Mann-Whitney tests on the null hypothesis of equality of samples between the two locations within the same country, or between the two countries.

S1.3 Identification of Nash equilibria and cooperative solution

The Collective Risk Social Dilemma (CRSD) is an *n*-person game (in our experiment *n=6*) where each player *i* is initially endowed with an equal amount of money *w* and can contribute some amount c_i, with $0 \le c_i \le w$, to a collective account in order to avoid a loss to his or her private account. If all players' total contribution, denoted by $C \equiv \sum_{j=1}^{n} c_j$, exceeds a given threshold *T*, there will be no loss for any player's private account of size $w - c_i$. If, however, $C < T$, there will be a loss of *L*% to each player's private account. We denote by $P = min\left\{\frac{C}{T}; 1\right\}$ the probability of loss avoidance (*PLA*). In case of loss, only a share $s = 1 - L$ of the private account will survive. The final payoff will be $w - c_i$ with probability *P* and $s \cdot (w - c_i)$ with probability $1 - P$.

The stage game with no sanctioning

First, we rule out the possibility of sanctioning, and for simplicity we assess the interaction as if it was played over just one period, rather than over ten periods. Even if it is plausible that many individuals dynamically conditioned their behavior on the observation of what others did in the previous period, the basic insights over the strategic nature of the interaction can be better seen considering a one-shot reduced form game. The expected payoff for a *risk neutral* player[2] with purely selfish preferences is then given by:

$$EU_i(c_i, c_{-i}) = [w - c_i] \cdot \frac{\min\{\sum_{j \neq i} c_j + c_i, T\}}{T} + s \cdot [w - c_i] \frac{T - \min\{\sum_{j \neq i} c_j + c_i, T\}}{T} \quad (1)$$

where *E* is the expected value operator, and $c_{-i} = (c_1, ..., c_{i-1}, c_{i+1}, ..., c_n)$ is the strategy profile of the other players except *i*.

To determine the non-cooperative equilibrium with $0 \le C \le T$ (thus omitting the Min-Operator), we differentiate $EU_i(c_i, c_{-i})$ with respect to c_i to obtain:

[2] A risk neutral player is indifferent between a lottery with uncertain payoffs and its expected value.

$$\frac{\partial EU_i(c_i,c_{-i})}{\partial c_i} = (1-s)\frac{w-c_i}{T} - \frac{\sum_{j\neq i} c_j + c_i}{T} - s\frac{T-\sum_{j\neq i} c_j - c_i}{T} = \frac{(1-s)\left(w-\sum_{j\neq i} c_j - 2c_i\right)}{T} - s \tag{2}$$

This term is negative for all strategy profiles (c_i, c_{-i}) whenever $T > w(1-s)/s$, i.e. if the threshold is sufficiently large.[3] Thus if

$$T \geq \frac{1-s}{s} w \tag{3}$$

contributing nothing is the unique symmetric non-cooperative Nash equilibrium with expected (= sure) payoff of $s \cdot w$ per player. On the other hand, if $T \leq nw/(s/(1-s)+n+1)$ then in the symmetric non-cooperative solution players contribute as much to avoid all risk. Finally, if $nw/(s/(1-s)+n+1) < T < (1-s)w/s$, there is an interior unique symmetric equilibrium with positive contributions given by

$$c_i^N = \frac{w - \frac{s}{1-s}T}{(n+1)}, \quad i = 1,...,n \tag{4}$$

It is interesting to note that both a higher threshold *T* and a higher survival rate *s* lead to lower equilibrium contributions, while more initial wealth and thus also a higher value at risk, increase contributions.

Cooperative solution

The cooperative solution maximizes the sum of individual expected payoffs:

$$E\sum_{i=1}^{n} U_i(c_1,...,c_n) = [nw-C]P + s[nw-C][1-P] \tag{5}$$

where *E* again is the expected value operator, and $C = \sum_{i=1}^{n} c_i$ denote total contributions. The first-order necessary condition for an interior solution is given by

$$\frac{1}{T}\left[nw(1-s) - sT - 2C(1-s)\right] = 0 \tag{6}$$

[3] It is easy to see that the second-order condition for a maximum is satisfied.

or solving for the total (group) contributions:

$$C^* = \frac{n \cdot w}{2} - \frac{sT}{2(1-s)} \tag{7}$$

An interior solution results if $C^* < T$. It is easy to see that this is equivalent to

$$T > \frac{nw(1-s)}{2-s} \tag{8}$$

while a corner solution $C^* = T$ by which the group eliminates all risk is optimal if

$$T \leq \frac{nw(1-s)}{2-s} \tag{9}$$

From (7) it is easy to see that $\partial C^* / \partial T < 0$, $\partial C^* / \partial w > 0$, and $\partial C^* / \partial s = -T/(2(s-1)^2) < 0$. Thus, both a higher threshold and a higher survival rate induce optimal contributions to fall, while higher wealth triggers more contributions since more is at risk.

Set of cooperative profiles:

Note that the optimal aggregate solution given by (7) in the interior case (i.e. positive contributions), and by $C^* = T$ in the corner case, can be generated by many different contribution profiles $\vec{c}^* = (c_1^*, ..., c_n^*)$ with $\sum_{i=1}^{n} c_i^* = C^*$ and $\vec{c}^*$ satisfying the individual rationality constraint for each player, i.e. no player is worse off as in the non-cooperative equilibrium:

$$EU_i(c_i^*, c_{-i}^*) \geq EU_i(c_1^N, ..., c_n^N) \tag{10}$$

If (3) holds, i.e. no player wants to contribute in equilibrium, (10) is satisfied if

$$c_i^* \leq \overline{c} \equiv w\frac{(1-s)C^*}{s \cdot T + (1-s)C^*} \tag{11}$$

Thus, all strategy profiles $\vec{c}^* = (c_1^*, ..., c_n^*)$ satisfying $\sum_{i=1}^{n} c_i^* = C^*$ and (11) are cooperative and individually rational outcomes.

Risk aversion

The above analysis was based on the hypothesis that players are risk neutral, i.e. indifferent between a lottery with uncertain payoffs and its expected value for sure. People who prefer the certain amount to a lottery whose expected payoff equals that amount are said to be risk averse, while people with opposite preferences are called risk seekers or risk lovers. Risk averse preferences can be introduced through a concave – rather than linear – objective function (or utility function) $U(x)$ with $U'(x) > 0$ and $U''(x) < 0$ (in the risk neutral case $U(x)$ being linear), defined over the space of money amounts x. In this case the objective function is:

$$EU(c_i, c_{-i}) = U(w - c_i) \cdot P + U([w - c_i]s) \cdot (1 - P) \tag{12}$$

Similar to the risk neutral case, one can show that contributing nothing is the only equilibrium if the threshold for avoiding any loss, *T*, is sufficiently high, and a unique interior symmetric equilibrium exists if *T* is sufficiently low. For a utility function of the form $U(x) = x^a$ with $0 < a < 1$ – referred to as constant relative risk aversion – equilibrium conditions are given by:

$$c_i^N = \frac{w - \dfrac{as^a}{1 - s^a} T}{(an + 1)}, \quad i = 1, \ldots, n \tag{13}$$

whereas the cooperative total contributions are determined by

$$C^* = \frac{nw}{(1 + a)} - \frac{as^a T}{(1 + a)(1 - s^a)} \tag{14}$$

Nash Equilibrium private contributions are non-decreasing in *a*, i.e., more risk aversion (a lower *a*) leads to lower contributions for an interior equilibrium. This result, counter-intuitive at first glance, is due to the fact that a share *s* of the private account will be earned with certainty even if the loss event occurs. The optimal individual response to others' contribution in the Nash Equilibrium will balance the marginal benefit of increasing the *PLA* and the marginal cost of reducing the amount that is earned even if the loss event occurs. Since the impact of increasing the *PLA* by contributing is relatively low from the individual standpoint, more risk-averse individuals will prefer to allocate

more money towards increasing the amount that is earned even in the case of loss. Hence, they will contribute *less* to the group account.[4]

By contrast, for an interior cooperative solution, total contributions are decreasing in *a*, which means that they are increasing if players get more risk-averse. The reason is that the cooperative solution internalizes the positive externality that contributing to the group account has on increasing *P* for everyone. From the group perspective, increasing *P* reduces the possibility of loss for *n* individuals, and this effect now dominates the effect of increasing the money that is earned if the loss event occurs for *n* individuals. Therefore, from the group perspective, increased risk-aversion by group-members will lead to *more* contribution to increase *P*.

The stage game with sanctioning

The above analysis can be extended to the case of sanctioning. In this case, personal accounts are equal to: $\sum_{t=1}^{\tau}(w_t - c_{it} - d_{it} - e_{it})$, where d_{it} is the sum of tokens spent by individual *i* in period *t* to deduct tokens from other players, while e_{it} is the number of tokens deducted from the account of individual *i* in period *t* as a result of sanctions by other players. With respect to the NE, agents willing to maximize their expected payoffs will not punish others, because this is costly to them and, according to the NE, players should continue to contribute nothing even when being sanctioned. As for the CS, there is no need to sanction, because players already achieve the course of action that maximizes total payoffs. Therefore, both the NE and CS with sanctions coincide with the NE and CS without sanctions.

It is clear that the NE does not take into account other motivations that individuals may have, such as a desire to pursue the group interests, altruism, concerns for efficiency, and reciprocity. It is nonetheless customary in economics to use the NE as a benchmark theoretical solution to analyze the strategic outcomes if people are only concerned with the maximization of their individual payoffs.

[4] Note that, when risk aversion is sufficiently high, (i.e., *a* is low), the marginal value of money when c=0 is very high, so the equilibrium contribution is zero.

S1.4 Comparison of international Open, international Blind, and national treatments

S1.4.1 Comparison of cooperation

We reported analyses of the international Open treatment (O-treatment) in the main paper. In this section of the Electronic Supplementary Materials (ESM), we report the analyses concerning the international Blind treatment. Generally speaking, we find no significant differences between Open and Blind treatments in terms of contributions to the group account. However, we do find some differences in sanctioning.

First of all, as reported in the paper, the Probability of Loss Avoidance (PLA) achieved in the international Open treatments was very close in size - and not statistically significantly different - from the PLA in the international Blind treatments, particularly so in the NS-treatments (d=0.04; p=0.99; N=32; d is Cohen's d; all tests are two-sided Wilcoxon-Mann-Whitney (WMW) tests unless otherwise indicated), but also in the Sanction treatments (d=0.41; p=0.30; N=32).

As in the International Open treatments, we found no significant difference in the PLA between International Blind treatments and the German national treatments, both in the NS-treatment (d= 0.008; p= 0.93, N=32) and in the S-treatment (d= 0.003; p= 0.96, N=32). In the NS-treatment, the PLA in the international Blind treatments was higher than in the Russian national treatment, the effect size being medium, but insignificant at conventional levels (d= 0.72; p= 0.063, N=32). In the S-treatment, the PLA in the international Blind treatments was higher than in the Russian national treatment, with effect size being twice as large compared to the NS-treatment, and the null hypothesis of equality of distributions being soundly rejected (d=1.41; p= 0.0002, N=32). Moreover, the PLA was significantly higher in the S-treatment than in the NS-treatment in Blind

international treatments (d=1.38; p= 0.0026, N=32). Therefore, we observe the same pattern of differences between international and national treatments, and in the effect of introducing sanctions, in the Open and the Blind treatments.

Even in the Blind international treatments, the behaviour of German and Russian participants was indistinguishable from each other, both in the NS-treatment (d=-0.03; signrank WMW: p=0.90, N=16) and in the S-treatment (d=-0.02; p=1.00; N=16), using two-tailed WMW matched-pairs signrank tests (Fig. 3, Panel A). Moreover, Germans behaved not significantly differently in Blind and Open treatments in both the NS-treatment (d=-0.06; p=0.96; N=32) and the S-treatment (d=0.28; p=0.42; N=32). The same was true for Russians in both the NS-treatment (d=0.14; p=0.79; N=32) and the S-treatment (d=0.41; p=0.27; N=32).

Table S9 reports the results of non-parametric tests period by period over the null hypothesis that contribution by Russian participants and German participants in international treatments were different from contribution in the respective national treatments. Qualitatively, we observe the same pattern in the Blind and Open treatments.

As far as Russian participants are concerned, when sanctions were not available, the null hypothesis of equality of distributions between national and international treatments was rejected, at conventional levels, only in period 1 (d=-0.76; p=0.027; N=32) and period 8 (d=-0.82; p=0.033; N=32) in the Blind treatment. The null was rejected only in two periods as well - period 7 (d=-1.15; p=0.0042; N=32) and period 8 (d=-0.82; p=0.011; N=32) - in the comparison of the Open NS-treatment and the national treatment. When sanctions were available, the null hypothesis of equality of distributions between national and the international Blind treatment was rejected in all periods after the second at significance levels of $p<0.001$. The same was true in the comparison between the Open and the national treatment, albeit at somewhat lower significance levels, and

with the exception of Period 4 where the null was not rejected at conventional levels (p=0.096; Table S9). Moreover, non-parametric tests failed to reject the hypothesis of equality of distribution by Russian participants in the Blind and Open treatments in the NS-treatment for any period of interaction (p-values in analyses codes, not reported here). In the S-treatment, the null was only rejected in period 3 (d=0.88; p=0.043; N=32) and period 4 (d=0.69; p=0.045; N=32).

As far as German participants are concerned, no difference was observed, at conventional levels, between contributions in the open treatments - either Blind or Open - and the corresponding national treatments. Likewise, the null of equality of distribution in contributions by German participants in the Open and Blind international treatments was not rejected in any period of interaction (p-values in analyses codes, not reported here). Overall, we can conclude that cooperation in the Open treatment was indistinguishable from cooperation in the Blind treatment for both Russian and German participants.

S1.4.2 Comparison of sanctioning

Consistently with the analysis of cooperation, we considered each group as an independent observation. Russians spent about 68% on sanctions more than Germans in national treatments (Fig. S8), the difference being statistically insignificant although the effect size was medium (d=0.58; p=0.15, N=32).

German participants spent on sanctions more in the Open-treatment than in the Blind treatment –the difference being statistically insignificant, at conventional levels, but the effect size being medium (d=0.59; p=0.052, N=32). The amount spent on sanctions in the international O-treatment is significantly higher than the amount spent in the national treatment (d=0.70; p=0.021, N=32), while no significant difference is found between the Blind international treatment and the national treatment (d=0.13; p=0.60,

N=32). No appreciable difference across treatments was found for Russian participants (see analyses codes).

S1.4.3 Comparison of antisocial and prosocial sanctioning

We defined anti-social sanctioning (*AS*) as an *ego* punishing an *alter* having contributed no less than the group median. An alternative definition used in the literature identifies *AS* as an *ego* punishing an *alter* having contributed no less than *ego* [20,25,56]. Pro-social sanctioning (*PS*) is defined as the residual category of *AS*, i.e. sanctioning targeting either *alters* who are contributing less than the median in the first definition, or alters having contributed less than *ego* in the alternative definition. Results are qualitatively equivalent using either definition (analyses not reported, available upon request). Previous studies observed significantly higher levels of *AS* in Russia than in Germany [19,20,28]. Consistently with the analysis of cooperation, we considered each group as an independent observation. We constructed the mean of *AS* and *PS* for each group (or (sub)group of participants from the same nationality within a group) dividing the total number of tokens spent for either *AS* or *PS* in a (sub)group by the number of people making up a (sub)group, that is, six people for the national treatments and three people for the international treatments.

Fig. S9 reports average *AS* and *PS* across treatments and nationality. Russians spent on average 2.52 times as much as Germans for *AS* in national treatments, although the difference is not significant at conventional levels (d= -0.84; p=0.055, N=32; Fig. S9). Average *AS* was 2.2 tokens in Russian national treatments (out of 100 tokens overall available individually for sanctioning over the 10 rounds), and 0.86 tokens in German national treatments. This difference has a large effect size but is at the margin of statistical significance (d=0.84; p=0.055, N=32). Russians spent more than Germans for *PS* in national treatments, too, but differences were smaller (d=0.45; p=0.21, N=32). The relatively modest amount spent in *AS* compared to other experiments is arguably caused by the possibility to identify who sanctioned others, a characteristic that has been proved to reduce sanctioning –especially *AS* – for fear of retaliation [25].

While the patterns of Germans and Russians' sanctioning involved in International B-treatments tended to be similar to what observed in national treatments (see analyses

codes), some differences emerged in the International Open treatments. Germans significantly increased the amount of *PS* in the International Open treatment (6.1 tokens) compared to the German national treatment (3.9 tokens) (d=0.60; p=0.044, N=32), while the difference in *AS* was insignificant (d=0.40; p=0.19, N=32). Russians increased their level of PS in the International O-treatment (8.4 tokens) compared to the Russian national treatment (5.8 tokens), the difference having a small to medium statistically insignificant effect (d=0.43;p=0.34;N=32). Notably, Russians nearly halved the amount of AS in the International O-treatment (1.2 tokens) compared to the Russian national treatment (2.2 tokens), the difference having a medium but statistically insignificant effect (d=0.62; p=0.15, N=32).

It is also noteworthy that in the International B-treatment Russians' AS was significantly higher than Germans (d= 0.47; two-tailed WMW matched-pairs signrank test: p=0.042, N=16), while any difference in AS between Russians and Germans all but disappeared in the International O-treatment (d= 0.23; two-tailed WMW matched-pairs signrank test: p=0.95, N=16).

As for PS sanctioning, there was no significant difference between Russian and German participants in either the International Blind treatment (d= 0.35; two-tailed WMW matched-pairs sign rank test: p=0.24, N=16) or the International Open treatment (d= 0.40; two-tailed WMW matched-pairs sign rank test: p=0.44, N=16).

S1.5 Analysis of the impact of sanctions on contributions

We analysed the capacity of sanctions to increase cooperation through an OLS estimator of an econometric model using as dependent variable the difference in Contribution to the collective fund between the current period and the previous period – i.e. $\Delta Contribution = Contribution_t - Contribution_{t-1}$. Even if the data have a panel structure, individual random effects are obliterated by the fact that the dependent variable is a difference of individual-level variables. $Sanction\_Loss_{t-1}$ is the key independent variable in the analysis reported in Table S11. It is the number of tokens being deducted from a participant's personal account in the previous period because of sanctioning by other group members. $Sanction\_Loss_{t-1}$ can range from 0 to 55 tokens (Table S6). Fig. S7 reports the distribution of $Sanction\_Loss_{t-1}$ by treatment. The model in Table S11, column 1, includes fixed effects for treatments - RUS_NAT_S being the omitted category - and for periods. Given that treatments were randomly assigned to groups of participants,

one possibility is to cluster standard errors at the group level to obtain standard errors robust to heteroskedasticity [57]. Nevertheless, we follow the more conservative approach [58] suggesting to consider different levels of clustering – individual, group, and session level in our case – and then choose the level of clustering associated with the lowest average within-cluster standard deviation, which yields the highest possible standard error correction for heteroskedasticity. By construction, this approach minimises the possibility of incurring in false-positive treatment effects, i.e. accepting that a treatment effect exists when this is not the case. In our case, the mean standard deviations for $\Delta Contribution$ 8.08 for the individual level, 9.03 for the group level, and 9.30 for the session level. We then opt for clustering standard errors at the individual level [59]. This model was reported in Table 2 and commented in the main paper.

The model in Table S11, column 2 adds demographic characteristics that are "exogenous" to the participant – namely, country of birth, age, gender, and parents' education. The latter variable is modelled as a pair of dummy variables identifying whether one or both parents have attained a university degree, neither parent holding higher education being the omitted category. The model of column 3 adds demographic variables that are, at least partly, the result of the participant's decisions. Such are the participant's university degree – grouped into Humanities and Social Sciences, Mathematics and Natural Sciences, Economics, other disciplines being the omitted category – marital status, having participated in a university exchange program, religion – coded as Christian denomination in Germany (catholic and protestant), Christian denomination in Russia (orthodox), other denominations (Buddhists, Muslims, Orthodox in Germany, Protestant and Catholic in Russia), an index of environmental action, and a risk tolerance measure (see Section S6: Question 30). The index of environmental action is the first principal component of four questions asking whether participants buy environmentally-friendly goods, save water, participate in ecological movements, and are active in recycling (see Section S6: Questions 19-22, and section S1.7, Fig. S12 and S13). It is worth noting that the coefficient for $Sanction\_Loss_{t-1}$ remains stable to the inclusion of such demographic factors. We also note that men increased cooperation significantly less than women for every token of sanctioning (b=-0.54; p=0.007; Table S11, column 2) and this effect is robust to the inclusion of additional individual controls (b=-0.50; p=0.012, Table S11, column 3). The risk tolerance measure has a significant negative effect on the dependent variable (b=-0.14; p=0.004, Table S11, column 3). This

entails that less risk tolerant individuals increased their cooperation more than others after having been sanctioned. Economics students tended to be less reactive to sanctions than students in the residual category, although the variable was not significant at conventional levels (p=0.083, Table S11, column 3). No other individual control variable was statistically significant.

In models 4 – 6 we replicated the models in (1 – 3) adding the interaction terms between $Sanction\_Loss_{t-1}$ and the treatment dummies. The model in column 4 provides the coefficients reported in Table 2, columns 2–5, of the main paper, relative to the impact of $Sanction\_Loss_{t-1}$ in each treatment. t-tests on the null hypothesis of equality between treatment coefficients from the same model are reported in Table 2, columns 3–5. The introduction of demographic variables in columns 5-6 of Table S11 leaves the key interaction coefficients approximately unchanged, showing the robustness of the results to demographic characteristics. It is also worth noting the negative sign of all Period coefficients, and the sizable and highly significant coefficients for Period 9 and 10 ($p<0.001$ for either variable in all models in Table S11), Period 2 being the omitted category. This is the consequence of a decreasing trend in contributions across periods, with a markedly pronounced drop in contributions in the last two periods of interaction (Fig. S5). Nevertheless, the disciplining power of sanctioning did not seem to vary over time. Adding an interaction term between $Sanction\_Loss_{t-1}$ and the variable *Period* indicating the period of interaction returns an insignificant effect (p=0.41; regression not reported; see analysis codes, line 207 and following).

Table S12 replicates the above analysis using $Sanction_{t-1}$ instead of $Sanction\_Loss_{t-1}$ as the key independent variable to study the impact of sanctioning on next period contributions. $Sanction_{t-1}$ is defined as follows:

$$Sanction_{t-1} = \{\, 1 \quad if\ Sanction\_Loss_{t-1} > 0\ 0 \quad if\ Sanction\_Loss_{t-1} = 0 \,\} \tag{15}$$

Comparing the results from the models of Table S11 and Table S12 enables us to study whether the size of sanctioning was relevant in addition to and beyond the mere fact of having been sanctioned. Although the variety of motivations behind sanctioning is large [60], sanctions typically transmit the information that others are dissatisfied with an individual's past behaviour, particularly for failing to comply with injunctive norms as

perceived by other individuals in the group. For this reason, sanctioning transmits relevant information to the sanctioned individual in addition to the size of the sanctions. On average across treatments, being sanctioned increased cooperation by 4.3 tokens in the next period, compared with not being sanctioned (p<0.001, Table S12, column 1). The impact was significantly different from 0 in all treatments (p<0.001 for all of them, Table S12, column 4), was largest in GER_NAT_S (b= 4.89) and smallest in RUS_NAT_S (b= 3.97). This result suggests that sanctions did not need to be large to urge individuals to cooperate more. This intuition was also supported by the observation that a dummy variable identifying received sanctions of just one token significantly increased contribution in the next round (b= 1.81, p=0.003) in comparison with not being sanctioned. Conversely, a dummy variable identifying all sanctions larger than one token had a larger impact (b=5.04, p<0.001). The different impact of small and large sanctions on $\Delta Contribution$ is statistically significant (b=3.22, p<0.001), demonstrating that the size of sanctions also mattered in addition to receiving an almost "symbolic" sanction of one token. We did not find significant differences of $Sanction_{t-1}$ across treatments.[5] Demographic effects for $Sanction_{t-1}$ are similar to those observed for $Sanction\_Loss_{t-1}$. We conclude that sanctions spurred individuals to increase cooperation even when we do not consider the actual size of sanctions, suggesting that even small sanctions had a significant effect in increasing cooperation. Since we do not observe treatment differences in the way $Sanction_{t-1}$ affects $\Delta Contribution$, it was arguably the way sanctioned individuals reacted to relatively large sanctions that caused significant treatment differences in $Sanction\_Loss_{t-1}$.

S1.6 Econometric analysis of sanctions

We report results of econometric analysis to explain determinants of the decisions to sanction in Table S13. The dependent variable is the amount of sanction expenditure by a group member directed to another group member in each period. The dependent variable is discrete and ranges from 0 to 2 tokens. A Poisson regression is appropriate to study this variable. A negative binomial regression would give qualitatively similar results. We applied random effects at the level of each pair formed by a sanctioning agent and a sanctioned agent, and we clustered standard errors at the same level.

[5] p-values of pairwise tests on the null hypothesis that coefficients are equal to each other range from p=0.36 for the test involving GER_NAT_S to INT_B_S and p=0.91 for the test involving RUS_NAT_S and INT_B_S.

The first specification only included treatment and period dummies (Table S13, column 1). In international interactions, treatments were interacted with the participant's nationality. Sanctions were significantly higher in Russian national interactions than German national interactions (b= -0.51; p<0.001). It is noteworthy that Germans sanctioned significantly more when involved in the international Open treatment than in national interaction (b= 0.49; p=0.001), while no significant difference was found for Russian participants. Interestingly, Russian participants sanctioned significantly more than German participants in the International Blind treatment (b= 0.46; p=0.008), but not in the International Open treatment (b= 0.21; p=0.18).

The second model introduces *Counterpart Contribution* - the amount of contribution by the participant's counterpart - arguably an important determinant of the propensity to sanction. Controlling for this variable is particularly important to assess differences in the national treatments, because participants were faced with substantially different cooperation levels. *Counterpart Contribution* is indeed always a highly significant predictor of sanctioning (p<0.001, Table S13, columns 2-4). The sign is negative, which means, as expected, that higher cooperation attracted lower sanctioning. It is noteworthy that the introduction of this variable considerably affects the treatment effects examined before. Germans sanctioning rate in national treatment is no longer different from Russian sanctioning rate, and the sign is even positive (b=0.04; p= 0.78). This suggests that the higher sanctioning observed in the Russian National treatment in comparison with the German National treatment is not caused by higher intrinsic propensity by Russians to sanction others in comparison to Germans, but rather by the fact that Russians are faced with low cooperators with higher frequency. Germans sanctions significantly more in International Open treatment than in the National treatment even after controlling for *Counterpart Contribution* (b=0.48; p= 0.001). The same is true for Russians both in the International Blind treatment (b=0.67; p<0.001) and the International Open treatment (b=0.64; p<0.001).

These results changed only marginally after the introduction of demographic factors (Table S13, columns 3-4). Once again, gender and risk tolerance were significant predictors of behaviour. Men sanctioned significantly more than women (b=0.59; p<0.001; Table S13, column 4) and more risk tolerant individuals sanctioned significantly less than others (b=-0.06; p=0.012; Table S13, column 4).

We fitted the four additional models to the analysis of Antisocial Sanctioning (Table S13, column 5-8) and Prosocial Sanctioning (Table S13, column 9-12). We did this by restricting observations to those in which the counterpart had contributed above or at the same level as the group median to identify potential Antisocial Sanctioning. Likewise, we restricted observations to those in which the counterpart had contributed below the group median to identify potential Prosocial Sanctioning. The dependent variable is again the number of tokens assigned in sanctioning to a given counterpart, ranging from 0 to 2 tokens.

We commented on the key results for these models in the main paper. Here we only note that the lower propensity by Germans compared to Russians to sanction antisocially also holds without controlling for *Counterpart Contribution* (b=-0.97; p<0.001; Table S13, column 5). On the contrary, the higher propensity by Germans compared to Russians to sanction prosocially does not hold without controlling for *Counterpart Contribution* (b=-0.17; p=0.17; Table S13, column 9). Among demographic effects, men sanction significantly more than women both antisocially (b=0.49; p=0.002; Table S13, column 5) and prosocially (b=0.63; p<0.001; Table S13, column 12), while risk tolerance only predicts prosocial sanctioning (b=-0.07; p=0.001; Table S13, column 12), but not antisocial sanctioning (b=0.01; p=0.77; Table S13, column 5).

S1.7 Analysis of contributions

We now look at determinants of contribution. We take as the dependent variable the total number of tokens contributed to the group account over the 10 periods of interaction (*Total Contribution*). We fit an OLS estimator, with the set of covariates being the same as that used for the model described in Table S11 – except for the exclusion of past sanctions. Treatment fixed effects now include NS treatments, RUS_NAT_S being the omitted category. We follow the approach described in Section S1.5 and apply clustering of standard errors at the group level, because the average standard deviation of *Total Contribution* is lower at this level (s.d.= 84.7) than at the session level (s.d.= 89.5).

The results are reported in Table S14. We commented on the results concerning the Environmental Action Index (defined in Section S1.5) in the main paper. Here we note that among the demographic variables we included, only gender appears to be a significant predictor of *Total Contribution*. We estimate that men contributed about 21 fewer tokens than women (over a total of possible contributions of 500 tokens over the ten periods) (p= 0.001), see Table S14, columns 1 to 3. It is also noteworthy that more risk tolerant individuals contribute less than others (b=-5.16; p=0.001; Table S14, column 3).

S1.8 Analysis of payoffs

It is *a priori* not clear whether average expected payoffs in S-treatments should be higher than in NS-treatments. On the one hand, sanctioning others is costly, and this cost will lower payoffs in S-treatments. On the other hand, sanctioning induces higher cooperation levels, which reduce the risk of the loss event to occur and thus raise earnings in S-treatments compared to NS-treatments. Average expected individual payoffs were significantly higher in NS-treatments (263.9 tokens) than in S-treatments (246.8 tokens) ($d=1.03$; $p<0.0001$; $N=128$). This is also the case in every pairwise comparison of NS- and S-treatments in either National or International treatments (Fig. S11, analyses codes). By construction, the Cooperative Solution (CS) for risk-neutral agents maximises expected group payoffs. This implies that, as participants in the S-treatments contributed more than what is prescribed by the CS, they incurred a cost in comparison with the optimal contribution level. An interpretation of this result is that individuals are predominantly risk-averse and will thus collectively prefer a level of Average Probability of Loss Avoidance (PLA) such as Point C in Fig. 1 in the main paper. Nonetheless, such a high level of PLA was only achieved when sanctions were available. Without sanctions, it is plausible that, as posited by previous research [20,28], participants will withhold cooperation as a form of indirect punishment, thus lowering the PLA.

It has been demonstrated that the payoff difference between sanction and no-sanction treatments is sensitive to the length of the interactions. With a low number of interactions, payoffs tend to be higher in NS than S-treatments [61], while with longer interactions S-treatments outperforms NS-treatments. The reason is that it takes time for participants to realise that people are ready to sanction, thus sanctioning costs are reduced in the long run [62]. Since interactions were relatively few in our experiment, it could be the case that mean payoffs in the S-treatments would have been higher with longer interactions. Nonetheless, the apparent preference for insurance above the financially optimal level seems to entail that participants accepted to pay an extra cost for higher safety.

Finally, there was no significant difference in payoffs accruing to German and Russian participants considering all international treatments together ($d=-0.10$; WMW signrank: $p=0.93$, $N=64$). Neither was there a significant differences in expected payoffs

between Germans and Russians in the national treatments (d=0.39; WMW signrank: p=0.30, N=64).

S1.9 Generalizability of results to nationally representative samples

Using university students' samples is subject to several types of biases that could prevent generalizability of results (see section S4.5).

In order to further test the representativeness of our sample, we have conducted an econometric exercise to estimate the amount of bias that running our experiment with a student sample introduces in comparison to a general sample. We have constructed an – admittedly basic – econometric model, in which some variables from our post-experiment questionnaire are used to predict behaviour in the experiment. Such variables are gender, generalised trust in others, and the extent to which the participants see themselves as part of the local, national, and world community, as well as the construal of the self as an autonomous individual. These variables are potential predictors of cooperation. In particular generalised trust is normally positively associated with cooperation, while perceiving to be an autonomous individual is likely to be negatively related with cooperation [40,63]. Moreover, the level of identification with local, national, and world communities can be considered as a predictor of cooperation in international interactions. This set of questions was also asked in the waves of the World Value Survey (WVS) conducted in 2011 in Russia and in 2013 in Germany with representative samples of the population [64]. Descriptive statistics for these variables in our sample and in the WVS are reported in Table S15. We have used this model to predict contribution in our student sample for different sets of treatments and different nationalities. We have then conducted an out-of-sample estimation to evaluate the cooperation levels by a representative sample from the WVS followed by estimating treatment effects for the hypothetical WVS sample.

This analysis shows that sizable differences in trust and social identification exist between our student sample and national representative samples. In particular, the student sample is more trusting in general others than the WVS sample in Germany, while the opposite occurs in Russia. In Germany, students see themselves as autonomous individuals and part of the world of the community more often, and see themselves part of the local community and national community less often, than the national sample. In

Russia, students see themselves as autonomous individuals and part of the local community more often, and see themselves part of the national and world community less often, than the national sample. Sizable differences on these traits exist between the two countries, as Germans perceive themselves as autonomous individuals, and as members of the local and world communities more than what Russians do, while Russians perceive themselves as members of the national community more than Germans do.

In an econometric model including both countries and all treatments, Generalised Trust is the strongest predictor of cooperation (b=17.2, p=0.010, N=736), particularly so in the national treatments (b=34.84, p=0.002, N=377). In international treatments, identification with the world community has a positive, albeit insignificant, sign (b=4.35, p=0.30, N=359), while identification with the national community has a negative – and insignificant – sign (b=-5.61, p=0.24, N=359). Seeing oneself as an autonomous individual is negatively associated with cooperation (b=-3.78, p=0.30, N=736). Our out-of-sample estimation predicts that a representative sample of the German population would be overall less cooperative than our student sample (Cohen's d=0.48), whereas a Russian representative sample would be substantially more cooperative than the student sample (d = 0.61). According to our estimates, representative samples would be more cooperative in international interactions than in national interactions both in Germany (d=0.21) and, particularly so, in Russia (d=2.02). According to this exercise, international cooperation would then be beneficial in comparison with national cooperation even with a nationally representative sample. According to this out-of-sample estimation, Russians would be more cooperative than Germans in international interactions (d=0.52).

S.2 Supplementary Tables

	National			Blind			Open		
	Germany	Russia	Total	Germany	Russia	Total	Germany	Russia	Total
Other lab is in same country	94.7	91.8	93.2	78.1	84.8	81.6	1.1	2.3	1.7
Other lab is either in Russia (for Germans) or Germany (for Russians)	0	0	0	0	3.3	1.7	90.9	86.5	88.7
Other lab is abroad but not in Russia / Germany	0	1.7	0.9	11.0	7.6	9.2	2.3	3.4	2.8
Do not know	5.3	6.6	5.9	11.0	4.4	7.5	5.7	7.9	6.8
Respondents	171	182	353	82	92	174	88	89	177

Table S4 | Distribution of beliefs over location of the other city. Instructions in National treatments specified that participants from the other city with which they were interacting were from the same country as the participant's city of residence. In Blind treatments, it was only said that the other university was located in "another city", without specifying the country. In the Open treatments, both German and Russian participants were told that they were interacting with other participants from Russia and Germany, respectively. We note that the distribution of beliefs in B-treatments is considerably closer to that in the National treatments than in the O-treatments.

	Germany	Russia
Age	0.26	0.10
Male	0.65	0.32
Father Education	0.38	0.33
Mother Education	0.56	0.54
University Degree	0.45	0.17
Degree: Humanities & Social Sciences	0.72	0.74
Degree: Mathematics & Natural Sciences	0.74	0.88
Degree: Economics	0.95	0.03*
University_Exchange	0.62	0.47
Years of residence outside country	0.18	0.41
Married	0.99	0.99
Environmental Action Index	0.74	0.11
Religion	0.38	0.67
Risk Tolerance	0.78	0.70

Table S5 | Test of exogeneity of treatment. We report the p-values of Kruskall-Wallis tests of the null hypothesis of equality of samples across treatments for a set of demographic, social background, university experience, and other personal characteristics, within either country. Except for a significant effect of Economics degree in Russia, the null hypothesis is never rejected at conventional levels of significance ($p<0.10$) for the variables being considered.

Total number of Tokens spent to reduce one group member's personal account by the other five group members	Number of Tokens deducted from this group member's personal account
0	0
1	1
2	3
3	6
4	10
5	15
6	21
7	28
8	36
9	45
10	55

Table S6: Relationship between tokens spent on sanctions and tokens deducted from the sanctioned participant's personal account.

22

	Germany			Russia		
	z	p-value	N	z	p-value	N
GER_NAT_NS	-1.19	0.23	16			
GER_NAT_S	-0.26	0.80	16			
RUS_NAT_NS				-1.16	0.24	16
RUS_NAT_S				-1.13	0.25	16
INT_B_NS	-1.050	0.29	16	0.11	0.92	16
INT_Blind_S	-	-	-	0.63	0.53	16
INT_O_NS	-0.32	0.75	16	-1.79	0.074	16
INT_Open_S	-1.89	0.059	16	1.16	0.25	16

23 **Table S7 | Analysis of within-country location differences: Contribution.**

24

	Germany			Russia		
	z	p-value	N	z	p-value	N
GER_NAT_S	1.79	0.074	16			
RUS_NAT_S				0.86	0.39	16
INT_Blind_S	-	-	-	-1.00	0.32	16
INT_Open_S	0.53	0.60	16	0.89	0.37	16

25 **Table S8: | Analysis of within-country location differences: Sanction.**

26

			Round									
Country	Treatment		1	2	3	4	5	6	7	8	9	10
	National Vs. International Blind - S	z-statistic	0.79	0.91	0.62	0.23	-0.53	-0.57	-0.28	-1.13	-0.98	-0.74
		P-value	*0.43*	*0.36*	*0.53*	*0.82*	*0.60*	*0.57*	*0.78*	*0.26*	*0.33*	*0.46*
	National Vs. International Open - S	z-statistic	1.23	1.93†	1.85†	1.13	1.28	0.64	-0.36	-1.04	-0.74	-0.038
		P-value	*0.22*	*0.054*	*0.065*	*0.26*	*0.20*	*0.52*	*0.72*	*0.30*	*0.46*	*0.97*
Germany	National Vs. International Blind -	z-statistic										
	NoS		-0.43	0.38	0.11	0.25	0.17	-0.21	-0.53	0.19	-0.49	-0.62
		P-value	*0.66*	*0.71*	*0.91*	*0.81*	*0.87*	*0.84*	*0.60*	*0.85*	*0.62*	*0.53*
	National Vs. International Open -	z-statistic										
	NoS		-0.79	-0.06	-0.09	-0.85	0.75	-0.42	-0.89	0.09	-0.17	0.08
		P-value	*0.43*	*0.95*	*0.92*	*0.40*	*0.45*	*0.68*	*0.38*	*0.92*	*0.87*	*0.94*
	National Vs. International Blind - S	z-statistic	-1.38	-1.38	-3.93***	-3.32***	-3.26**	-2.83**	-2.75**	-3.51***	-3.21**	-2.89**
		P-value	*0.17*	*0.17*	*0.0001*	*0.0009*	*0.0011*	*0.0047*	*0.0059*	*0.0004*	*0.0013*	*0.0039*
	National Vs. International Open - S	z-statistic	-0.74	-1.09	-2.45*	-1.68†	-2.38*	-2.64**	-2.30*	-2.34*	-3.22**	-3.00**
		P-value	*0.46*	*0.27*	*0.014*	*0.094*	*0.018*	*0.0083*	*0.021*	*0.019*	*0.0013*	*0.0027*
Russia	National Vs. International Blind -	z-statistic										
	NoS		-2.21*	-1.60	-1.30	-1.64	-1.40	-0.79	-1.73†	-2.13*	-1.92†	-1.21
		P-value	*0.027*	*0.11*	*0.19*	*0.10*	*0.16*	*0.43*	*0.083*	*0.033*	*0.054*	*0.23*
	National Vs. International Open -	z-statistic										
	NoS		-0.11	-0.66	-1.23	-1.02	-0.96	-0.62	-2.87**	-2.53*	-1.30	-0.47

	P-value	0.91	0.51	0.22	0.31	0.34	0.53	0.0042	0.011	0.19	0.64

Table S9 |Analysis of differences in contribution levels between International and National treatments per period of interaction

The Table reports results of Wilcoxon Mann-Whitney (WMW) ranksum tests over the null hypothesis that contributions in International treatments come from the same distribution as contribution in within-country national treatments. Tests are broken down by participants' nationality. Blind and Open treatments for either German or Russian participants are compared with the corresponding national treatment with participants from the same nationality. Sanction (No-Sanction) treatments in international treatments are compared with Sanction (No-Sanction) treatments in national treatments. The Table reports the z-statistic of the WMW test and the p-value of the test. A negative value of the z-statistic entails that the distribution of the national treatment tends to be more skewed towards the left – that is, observations tend to have lower value – than in the international treatment. The analysis is conducted at the group level, hence we have 32 observations for each test. *=$p<0.001$; **= $p<0.01$; *= $p<0.05$; †=$p<0.1$.

35

Country	Treatment	Statistics	Treatment	
			NS	S
Germany	National	Mean	65.1	86.2
		St. Dev.	(19.0)	(10.7)
		N	16	16
	International (Blind and Open pooled)	Mean	65.6	84.8
		St. Dev.	(20.8)	(12.6)
		N	32	32
	Blind	Mean	65.0	86.6
		St. Dev.	(23.3)	(13.3)
		N	16	16
	Open	Mean	66.3	83.0
		St. Dev.	(18.8)	(12.1)
		N	16	16
Russia	National	Mean	50.9	64.1
		St. Dev.	(21.2)	(19.7)
		N	16	16
	International (Blind and Open pooled)	Mean	64.2	84.1
		St. Dev.	(19.6)	(12.9)
		N	32	32
	Blind	Mean	65.6	86.8
		St. Dev.	(19.5)	(10.8)
		N	16	16
	Open	Mean	62.8	81.5
		St. Dev.	(20.3)	(14.5)
		N	16	16

36 **Table S10: Decomposition of impact of Sanctions and**
37 **Internationalisation of interaction on cooperation**
38

DEPENDENT VARIABLE	$Contribution_t - Contribution_{t-1}$					
	(1)	(2)	(3)	(4)	(5)	(6)
INDEPENDENT VARIABLES						
$Sanction\_Loss_{t-1}$	0.42***	0.42***	0.42***	0.36***	0.36***	0.34***
	[0.10]	[0.09]	[0.09]	[0.07]	[0.07]	[0.06]
GER_NAT_S	0.56*	0.81*	0.85*	0.21	0.42	0.48
	[0.28]	[0.38]	[0.39]	[0.30]	[0.42]	[0.42]
INT_Blind_S	0.91**	0.96**	1.06***	0.44	0.44	0.53
	[0.28]	[0.32]	[0.31]	[0.33]	[0.38]	[0.37]
INT_Open_S	0.62*	0.72*	0.87**	0.76*	0.81*	0.91*
	[0.31]	[0.32]	[0.32]	[0.34]	[0.39]	[0.38]
Country		0.38	0.58		0.29	0.55
		[0.35]	[0.41]		[0.31]	[0.38]
Age		-0.01	0.01		-0.01	0.01
		[0.03]	[0.03]		[0.03]	[0.03]
Gender (male =1		-0.54**	-0.50*		-0.54**	-0.50*
		[0.20]	[0.20]		[0.21]	[0.21]
High education one parent		-0.21	-0.21		-0.29	-0.28
		[0.24]	[0.25]		[0.25]	[0.26]
High education both parents		-0.39	-0.20		-0.40	-0.22
		[0.27]	[0.27]		[0.25]	[0.25]
Degree: Humanities and Social Sciences			0.06			0.14
			[0.29]			[0.28]
Degree: Mathematics and Natural Sciences			-0.16			-0.09
			[0.29]			[0.28]
Degree: Economics			-0.56†			-0.51†
			[0.32]			[0.30]
Married			-1.25			-1.14
			[0.80]			[0.82]
Univ. exchange program			-0.38			-0.34
			[0.31]			[0.30]
Environmental action index			0.07			0.04
			[0.09]			[0.09]
Catholic			0.16			0.04
			[0.35]			[0.37]
Protestant			-0.59			-0.40
			[0.40]			[0.35]
Orthodox			0.14			0.01
			[0.29]			[0.30]
Other religion			0.52			0.51
			[0.33]			[0.33]
Risk Tolerance			-0.14**			-0.14**
			[0.05]			[0.05]
Period 3	-0.36	-0.22	-0.11	-0.31	-0.17	-0.05

	[0.73]	[0.73]	[0.74]	[0.73]	[0.73]	[0.74]
Period 4	-1.67**	-1.59*	-1.49*	-1.60**	-1.52*	-1.40*
	[0.62]	[0.62]	[0.63]	[0.61]	[0.62]	[0.63]
Period 5	-0.77	-0.73	-0.43	-0.65	-0.62	-0.32
	[0.59]	[0.59]	[0.60]	[0.58]	[0.58]	[0.59]
Period 6	-1.13†	-1.19†	-1.25†	-1.04†	-1.11†	-1.16†
	[0.63]	[0.63]	[0.65]	[0.63]	[0.63]	[0.64]
Period 7	-1.98**	-1.84**	-1.76**	-1.88**	-1.74**	-1.66*
	[0.64]	[0.65]	[0.67]	[0.64]	[0.65]	[0.66]
Period 8	-1.03†	-1.19†	-1.28*	-0.97	-1.13†	-1.21†
	[0.61]	[0.61]	[0.62]	[0.61]	[0.61]	[0.62]
Period 9	-2.74***	-2.55***	-2.51***	-2.72***	-2.54***	-2.48***
	[0.67]	[0.67]	[0.68]	[0.66]	[0.67]	[0.68]
Period 10	-6.47***	-6.48***	-6.30***	-6.41***	-6.43***	-6.25***
	[0.77]	[0.78]	[0.81]	[0.77]	[0.79]	[0.81]
GER_NAT_S × $Sanction\_Loss_{t-1}$				0.31**	0.28**	0.29**
				[0.11]	[0.10]	[0.10]
INT_Blind_S × $Sanction\_Loss_{t-1}$				0.36*	0.36*	0.38*
				[0.16]	[0.16]	[0.16]
INT_Open_S × $Sanction\_Loss_{t-1}$				-0.07	-0.07	-0.03
				[0.17]	[0.16]	[0.16]
Constant	0.08	0.35	0.61	0.14	0.53	0.85
	[0.52]	[0.91]	[0.99]	[0.50]	[0.93]	[1.00]
Observations	3,456	3,366	3,240	3,456	3,366	3,240
Number of participants	384	374	360	384	374	360
R^2_within	0.0887	0.0902	0.0898	0.0923	0.0937	0.0932
R^2_between	0.0436	0.0512	0.0859	0.0499	0.0576	0.0931
R^2_overall	0.0748	0.0767	0.0789	0.0806	0.0822	0.0842
Number of clusters	384	374	360	384	374	360

Table S11 | Econometric analysis of the impact of sanction loss and demographic characteristics on cooperation change. We fit an OLS estimator to a model having as dependent variable the variation in Contribution between period t and t-1, for t=2,…,10. See Section S1.5 for variables' description and further details on econometric specification. Heteroskedasticity-robust standard errors clustered at the individual level are in brackets. The model in column (1) has been used to compute the aggregate effect of past sanctioning on contribution change, reported in column (1) of Table 2 in the main paper. The model in column (4) has been used to determine the coefficients for the effect of past sanctioning on contribution change by treatment, reported in column (2) of Table of the main paper. Wald tests on the null hypothesis that such coefficients differ from each other have been reported in columns (3)-(5) of Table 2 in the main paper. See codes of statistical analyses for output of such Wald tests for models of Table S11, columns (4)-(6). *** p<0.001; ** p<0.01; * p<0.05; † p<0.10.

40

DEPENDENT VARIABLE	$Contribution_t - Contribution_{t-1}$					
	(1)	(2)	(3)	(4)	(5)	(6)
INDEPENDENT VARIABLES						
$Sanction_{t-1}$	4.30***	4.22***	4.26***	3.97***	3.79***	3.78***
	[0.38]	[0.38]	[0.39]	[0.73]	[0.72]	[0.73]
GER_NAT_S	0.82**	0.87*	0.97*	0.60	0.60	0.76
	[0.28]	[0.41]	[0.41]	[0.44]	[0.53]	[0.54]
INT_BLIND_S	0.84**	0.81*	0.95**	0.86†	0.79†	0.86†
	[0.30]	[0.34]	[0.33]	[0.44]	[0.47]	[0.48]
INT_OPEN_S	0.73*	0.76*	0.90**	0.51	0.43	0.50
	[0.29]	[0.33]	[0.33]	[0.47]	[0.51]	[0.52]
Country		0.12	0.44		0.09	0.43
		[0.33]	[0.40]		[0.33]	[0.40]
Age		-0.01	0.01		-0.01	0.01
		[0.03]	[0.03]		[0.03]	[0.03]
Gender (male =1)		-0.57**	-0.52*		-0.55**	-0.50*
		[0.21]	[0.21]		[0.21]	[0.21]
High education one parent		-0.25	-0.26		-0.24	-0.26
		[0.27]	[0.28]		[0.27]	[0.28]
High education both parents		-0.28	-0.09		-0.28	-0.10
		[0.27]	[0.27]		[0.27]	[0.27]
Degree: Humanities and Social Sciences			0.06			0.05
			[0.28]			[0.29]
Degree: Mathematics and Natural Sciences			-0.26			-0.27
			[0.30]			[0.30]
Degree: Economics			-0.58†			-0.59†
			[0.30]			[0.30]
Married			-1.34			-1.43
			[0.94]			[0.97]
Univ. exchange program			-0.24			-0.25
			[0.28]			[0.28]
Environmental action index			0.06			0.06
			[0.09]			[0.09]
Catholic			0.10			0.13
			[0.39]			[0.40]
Protestant			-0.48			-0.51
			[0.36]			[0.36]
Orthodox			-0.00			-0.01
			[0.31]			[0.31]
Other religion			0.49			0.47
			[0.36]			[0.36]

Risk Tolerance			-0.14**			-0.15**
			[0.05]			[0.05]
Period 3	-0.26	-0.12	-0.02	-0.24	-0.10	-0.01
	[0.73]	[0.73]	[0.75]	[0.73]	[0.73]	[0.75]
Period 4	-1.58**	-1.50*	-1.45*	-1.57**	-1.49*	-1.44*
	[0.61]	[0.61]	[0.62]	[0.61]	[0.61]	[0.62]
Period 5	-0.86	-0.78	-0.51	-0.85	-0.76	-0.50
	[0.59]	[0.59]	[0.59]	[0.59]	[0.59]	[0.60]
Period 6	-1.03	-1.07†	-1.15†	-1.02	-1.06†	-1.14†
	[0.63]	[0.63]	[0.64]	[0.63]	[0.63]	[0.64]
Period 7	-1.93**	-1.78**	-1.72**	-1.92**	-1.77**	-1.72**
	[0.64]	[0.65]	[0.67]	[0.64]	[0.65]	[0.66]
Period 8	-0.82	-0.96	-1.04†	-0.81	-0.95	-1.03†
	[0.62]	[0.61]	[0.63]	[0.61]	[0.61]	[0.63]
Period 9	-2.69***	-2.50***	-2.49***	-2.69***	-2.50***	-2.50***
	[0.66]	[0.67]	[0.68]	[0.67]	[0.67]	[0.68]
Period 10	-6.30***	-6.31***	-6.15***	-6.31***	-6.32***	-6.16***
	[0.76]	[0.78]	[0.81]	[0.76]	[0.78]	[0.80]
GER_NAT_S × $Sanction_{t-1}$				0.92	0.93	0.60
				[1.13]	[1.11]	[1.09]
INT_Blind_S × $Sanction_{t-1}$				-0.12	-0.03	0.19
				[1.04]	[1.03]	[1.06]
INT_Open_S × $Sanction_{t-1}$				0.68	0.94	1.18
				[1.04]	[1.04]	[1.08]
Constant	-0.69	-0.12	0.21	-0.58	0.13	0.42
	[0.52]	[0.96]	[1.04]	[0.58]	[1.02]	[1.08]
Observations	3,456	3,366	3,240	3,456	3,366	3,240
Number of participants	384	374	360	384	374	360
R^2_within	0.0926	0.0931	0.0916	0.0926	0.0932	0.0919
R^2_between	0.0430	0.0484	0.0894	0.0446	0.0498	0.0890
R^2_overall	0.0795	0.0804	0.0827	0.0799	0.0810	0.0832
Number of clusters	384	374	360	384	374	360

Table S12 | Econometric analysis of the impact of sanction and demographic characteristics on cooperation change. The models replicate the analysis of Table S11 replacing $Sanction\_Loss_{t-1}$ with $Sanction_{t-1}$. The latter is a dummy variable identifying whether a participant had been sanctioned in the previous period, regardless of the sanction amount. See Section S1.5 for description of model and variables. *** p<0.001, ** p<0.01, * p<0.05, † p<0.10.

41

DEPENDENT VARIABLE	*Total sanction*				*Antisocial Sanctions*				*Prosocial Sanctions*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
INDEPENDENT VARIABLES												
Counterpart Contribution		-0.07***	-0.07***	-0.07***		-0.02**	-0.02**	-0.03**		-0.06***	-0.06***	-0.06***
		[0.00]	[0.00]	[0.00]		[0.01]	[0.01]	[0.01]		[0.00]	[0.00]	[0.00]
GER_NAT_S	-0.52***	0.04	0.02	0.40*	-0.97***	-0.81**	-0.88***	-0.58†	-0.17	0.32*	0.35*	0.75***
	[0.14]	[0.16]	[0.16]	[0.17]	[0.25]	[0.26]	[0.25]	[0.31]	[0.12]	[0.13]	[0.14]	[0.16]
INT_Blind_S	-0.40*	0.22	0.28	0.66***	-0.63*	-0.44	-0.35	0.05	-0.07	0.42**	0.45**	0.79***
	[0.16]	[0.18]	[0.18]	[0.19]	[0.27]	[0.29]	[0.30]	[0.32]	[0.15]	[0.16]	[0.17]	[0.19]
INT_Open_S	-0.03	0.52***	0.69***	1.13***	-0.46*	-0.27	-0.04	0.39	0.18	0.66***	0.83***	1.25***
	[0.14]	[0.16]	[0.16]	[0.18]	[0.23]	[0.25]	[0.27]	[0.33]	[0.13]	[0.15]	[0.15]	[0.18]
INT_Blind_S × RUS	0.46**	0.45**	0.52**	0.36†	0.93**	0.92**	0.93**	0.78*	0.33*	0.41*	0.47*	0.27
	[0.17]	[0.17]	[0.19]	[0.21]	[0.29]	[0.29]	[0.32]	[0.34]	[0.16]	[0.17]	[0.19]	[0.21]
INT_Open_S × RUS	0.21	0.12	-0.03	-0.41*	-0.19	-0.20	-0.45	-0.87*	0.24†	0.21	0.05	-0.35†
	[0.16]	[0.15]	[0.17]	[0.20]	[0.28]	[0.28]	[0.32]	[0.37]	[0.14]	[0.15]	[0.16]	[0.19]
Age			0.02†	0.00			0.02	0.00			0.03*	0.01
			[0.01]	[0.01]			[0.02]	[0.03]			[0.01]	[0.01]
Gender (male =1)			0.58***	0.59***			0.58***	0.49**			0.59***	0.63***
			[0.09]	[0.09]			[0.15]	[0.16]			[0.08]	[0.09]
High education one parent			0.14	0.13			0.13	0.01			0.19†	0.25*
			[0.12]	[0.12]			[0.22]	[0.23]			[0.11]	[0.12]
High education both parents			0.26*	0.22†			0.10	0.00			0.42***	0.40**
			[0.12]	[0.12]			[0.21]	[0.22]			[0.12]	[0.13]
Degree: Humanities and Social Sciences				0.03				-0.14				0.24†
				[0.15]				[0.25]				[0.14]

Degree: Mathematics and Natural Sciences				-0.20				-0.05				-0.07
				[0.15]				[0.25]				[0.15]
Degree: Economics				-0.05				-0.14				0.22
				[0.16]				[0.25]				[0.15]
Married				0.18				-0.52				0.20
				[0.28]				[0.51]				[0.28]
Univ. exchange program				0.52***				0.74**				0.38**
				[0.14]				[0.23]				[0.13]
Environmental action index				-0.01				-0.00				-0.01
				[0.04]				[0.06]				[0.04]
Catholic				-0.19				0.36				-0.44*
				[0.20]				[0.36]				[0.19]
Protestant				-0.06				-0.03				-0.21
				[0.16]				[0.31]				[0.16]
Orthodox				0.24†				0.24				0.13
				[0.13]				[0.22]				[0.13]
Other religion				-0.10				-0.03				-0.08
				[0.14]				[0.27]				[0.13]
Risk Tolerance				-0.06*				0.01				-0.07**
				[0.02]				[0.04]				[0.02]
Period 2	0.02	0.24**	0.25**	0.25**	0.32	0.29	0.30	0.33	0.01	0.24**	0.25**	0.25**
	[0.08]	[0.08]	[0.08]	[0.08]	[0.21]	[0.21]	[0.21]	[0.23]	[0.09]	[0.09]	[0.09]	[0.09]
Period 3	0.12	0.42***	0.42***	0.42***	0.47*	0.48*	0.50*	0.56*	0.15	0.43***	0.43***	0.42***
	[0.09]	[0.08]	[0.08]	[0.09]	[0.24]	[0.24]	[0.24]	[0.25]	[0.09]	[0.09]	[0.09]	[0.10]
Period 4	-0.08	0.28**	0.27**	0.27**	0.28	0.28	0.26	0.28	0.04	0.35***	0.35***	0.35***
	[0.09]	[0.09]	[0.09]	[0.10]	[0.23]	[0.23]	[0.23]	[0.25]	[0.10]	[0.10]	[0.10]	[0.10]
Period 5	-0.01	0.36***	0.35***	0.36***	0.06	0.05	0.06	0.14	0.14	0.49***	0.49***	0.49***
	[0.10]	[0.09]	[0.09]	[0.10]	[0.24]	[0.25]	[0.25]	[0.26]	[0.09]	[0.09]	[0.09]	[0.10]
Period 6	-0.05	0.34***	0.33***	0.33***	0.32	0.32	0.33	0.37	0.00	0.35***	0.35***	0.34**
	[0.10]	[0.09]	[0.09]	[0.10]	[0.22]	[0.22]	[0.22]	[0.25]	[0.10]	[0.10]	[0.10]	[0.11]

Period 7	-0.00	0.31***	0.31**	0.32***	0.12	0.08	0.05	0.20	0.07	0.38***	0.38***	0.37***
	[0.10]	[0.09]	[0.09]	[0.10]	[0.23]	[0.23]	[0.23]	[0.25]	[0.10]	[0.10]	[0.10]	[0.11]
Period 8	-0.04	0.34***	0.33***	0.35***	0.27	0.29	0.29	0.38	0.04	0.36***	0.35**	0.36**
	[0.10]	[0.10]	[0.09]	[0.10]	[0.22]	[0.22]	[0.22]	[0.24]	[0.10]	[0.11]	[0.11]	[0.11]
Period 9	-0.11	0.12	0.10	0.05	-0.04	-0.06	-0.06	-0.09	-0.05	0.18	0.15	0.12
	[0.11]	[0.10]	[0.10]	[0.11]	[0.26]	[0.26]	[0.26]	[0.29]	[0.11]	[0.11]	[0.11]	[0.12]
Period 10	0.45***	0.21*	0.19†	0.18†	0.54*	0.46†	0.43†	0.48†	0.38***	0.17†	0.17	0.18
	[0.09]	[0.10]	[0.10]	[0.11]	[0.24]	[0.24]	[0.24]	[0.27]	[0.10]	[0.10]	[0.10]	[0.11]
Constant	-1.88***	-0.87***	-1.92***	-1.37***	-2.80***	-2.17***	-2.99***	-2.95***	-1.52***	-1.00***	-2.18***	-1.70***
	[0.11]	[0.12]	[0.32]	[0.33]	[0.21]	[0.28]	[0.63]	[0.72]	[0.11]	[0.11]	[0.28]	[0.32]
Observations	19,200	19,200	18,700	18,000	11,990	11,990	11,663	11,240	7,210	7,210	7,037	6,760
Sanctioning opportunities per round	1,920	1,920	1,870	1,800	1,870	1,870	1,821	1,755	1,710	1,710	1,669	1,607
Chi2	2125	1939	1963	2405	1629	1535	1553	1818	1111	1070	1185	1343
Number of clusters	3840	3840	3840	3840	3840	3840	3840	3840	3840	3840	3840	3840
Ln alpha	1.38*	1.20†	1.13†	1.10†	2.29***	2.29***	2.25***	2.24***	0.95*	0.87*	0.76	0.74
	[0.54]	[0.62]	[0.64]	[0.65]	[0.61]	[0.62]	[0.63]	[0.64]	[0.40]	[0.43]	[0.47]	[0.49]

Table S13 | Econometric analysis of sanctioning in experiments. We fitted a Poisson regression with random effects at the level of individual-counterpart pair. The three dependent variables are Total Sanctions (Columns 1-4), Antisocial Sanctions (Columns 5-8), and Prosocial Sanctions (Columns 9-12). The dependent variable is again the number of tokens assigned in sanctioning to a given counterpart in a given period, and ranges from 0 to 2 tokens. While Total Sanctions refer to the whole set of possible sanction possibilities, we study Antisocial Sanctions by restricting observations to those in which the counterpart had contributed above or at the same level as the group median (Columns 5-8). Likewise, we identify potential Prosocial Sanctioning restricting observations to those in which the counterpart had contributed below the group median. The covariates are the same as those used in the models of Table S11 (except for the omission of *Sanction_Loss*$_{t-1}$) and include the Contribution by the recipient of the sanction (*Counterpart Contribution*). See Section S1.5 for description of variables and Section S1.6. for model description. Heteroskedasticity-robust standard errors clustered at the individual-counterpart level are in brackets. *** p<0.001, ** p<0.01, * p<0.05, † p<0.10.

DEPENDENT VARIABLE	Total Contribution					
	(1)	(2)	(3)	(4)	(5)	(6)
INDEPENDENT VARIABLES						
RUS_NAT_NS	-45.76†	-44.39†	-42.90†			
	[24.35]	[24.48]	[25.33]			
GER_NAT_S	74.55***	75.33***	76.83***	73.58**	73.88**	78.77***
	[21.31]	[21.23]	[21.34]	[21.81]	[21.62]	[21.01]
GER_NAT_NS	-0.81	-0.00	3.12			
	[24.23]	[24.13]	[24.48]			
INT_Blind_S	78.08***	78.77***	79.22***	77.41***	78.06***	81.24***
	[19.46]	[19.33]	[19.36]	[19.44]	[19.21]	[18.82]
INT_B_NS	2.56	3.29	4.12			
	[23.62]	[23.58]	[23.08]			
INT_Open_S	63.34**	62.87**	67.16***	62.24**	61.69**	70.10***
	[19.39]	[19.26]	[18.71]	[19.54]	[19.21]	[18.13]
INT_OP_NS	-1.39	-0.35	2.87			
	[22.63]	[22.61]	[23.15]			
Country	0.54	2.79	-3.09	0.90	3.67	13.88
	[8.93]	[9.03]	[11.39]	[10.26]	[10.42]	[11.42]
Age	1.20	1.03	0.38	1.22	0.99	0.69
	[0.84]	[0.85]	[1.00]	[0.83]	[0.83]	[0.97]
Gender (male =1	-21.37***	-20.62***	-18.40**	-3.05	-1.25	-0.05
	[6.06]	[6.06]	[6.57]	[6.88]	[7.01]	[7.60]
High education one parent	5.09	5.26	7.34	-4.54	-3.78	-1.22
	[7.91]	[7.92]	[7.96]	[7.07]	[6.88]	[7.67]
High education both parents	-0.93	-0.70	1.94	-2.37	-1.86	-0.80
	[7.75]	[7.69]	[7.75]	[9.55]	[9.24]	[8.49]
Environmental action index		3.10	3.18		4.93†	5.12†
		[2.32]	[2.40]		[2.77]	[2.96]
Degree: Humanities and Social Sciences			9.08			6.00
			[9.18]			[9.72]
Degree: Mathematics and Natural Sciences			6.40			5.34
			[8.96]			[8.04]
Degree: Economics			0.06			5.41
			[9.44]			[9.25]
Married			17.11			22.24
			[30.09]			[26.47]
Univ. exchange program			6.56			-11.70
			[9.56]			[10.95]
Catholic			-11.60			-0.31
			[8.89]			[10.09]
Protestant			-9.74			-6.36
			[11.11]			[13.22]
Orthodox			6.35			-12.60
			[11.06]			[11.66]
Other religion			-8.61			-6.37
			[9.93]			[12.53]

Risk Tolerance			-5.16***			-5.88**
			[1.51]			[1.94]
Constant	208.22***	209.22***	248.86***	202.79***	204.05***	234.34***
	[28.34]	[28.27]	[30.94]	[26.39]	[26.01]	[27.00]
Observations	746	744	725	374	372	360
R^2	0.22	0.22	0.24	0.20	0.21	0.26
Number of clusters	128	128	128	64	64	64

Table S14 | Econometric analysis of total individual cooperation in the experiment. We fit an OLS estimator having total individual contributions (Total Contribution) as the dependent variable. The set of covariates is the same used for the model described in Table S11 – except for past sanctions. Heteroskedasticity-robust standard errors clustered at the group level are in brackets. See Section S1.5 for variable description and Section S1.7. for model description. *** p<0.001; ** p<0.01; * p<0.05; † p<0.10.

52

Country / Source	Male	Individual Citizen	Local Citizen	Country Citizen	World Citizen	Trust
GER-Exp						
Mean	0.49	2.42	1.41	1.87	2.16	0.62
Median	0.50	0.71	0.95	0.87	0.76	0.49
Min	0	0	0	0	0	0
Max	1	3	3	3	3	1
N	384	383	384	384	383	369
GER-WVS						
Mean	0.50	2.18	2.13	2.25	1.69	0.58
Median	0.50	0.84	0.83	0.73	0.94	0.49
Min	0	0	0	0	0	0
Max	1	3	3	3	3	1
N	2046	1978	2025	2017	1976	2017
RUS-Exp						
Mean	0.47	1.10	1.31	2.15	0.98	0.45
Median	0.50	0.96	1.01	0.93	0.95	0.50
Min	0	0	0	0	0	0
Max	1	3	3	3	3	1
N	384	384	384	372	384	372
RUS-WVS						
Mean	0.45	0.93	0.89	2.58	1.49	0.71
Median	0.50	0.98	0.95	0.67	1.00	0.45
Min	0	0	0	0	0	0
Max	1	3	3	3	3	1
N	2500	2015	2164	2448	2277	2350

Table S15 | Analysis of differences in gender and social identity in our sample and World Value Survey sample. Descriptive statistics of variables that were used in both our post-experiment questionnaire and in the World Value Survey (WVS) in Germany and Russia are reported. GER-Exp and RUS-Exp denote data from our own study, while GER-WVS and RUS-WVS denote data from the World Value Survey. Male is a dummy variable identifying males. The other variables are answers to the Question 25 in the questionnaire, which asked participants to express their agreement with the following statements: "I see myself as an autonomous individual" (for 'Individual Citizen'); "I see myself as part of my local community" (for 'Local Citizen'); "I see myself as part of the Russian (for Russian version) / German (for German version) nation." (for 'Country Citizen'); "I see myself as a world citizen." (for 'World Citizen'). Answers were given on the following scale: 0 = "Strongly disagree"; 1 = "Disagree"; 2= "Agree", 3= "Strongly agree". (The original scale was reversed, see Question 25). These questions were also asked in the 2011 WVS wave conducted in Russia and in 2013 WVS wave in Germany, thus making a comparison possible.

S.3 Supplementary Figures

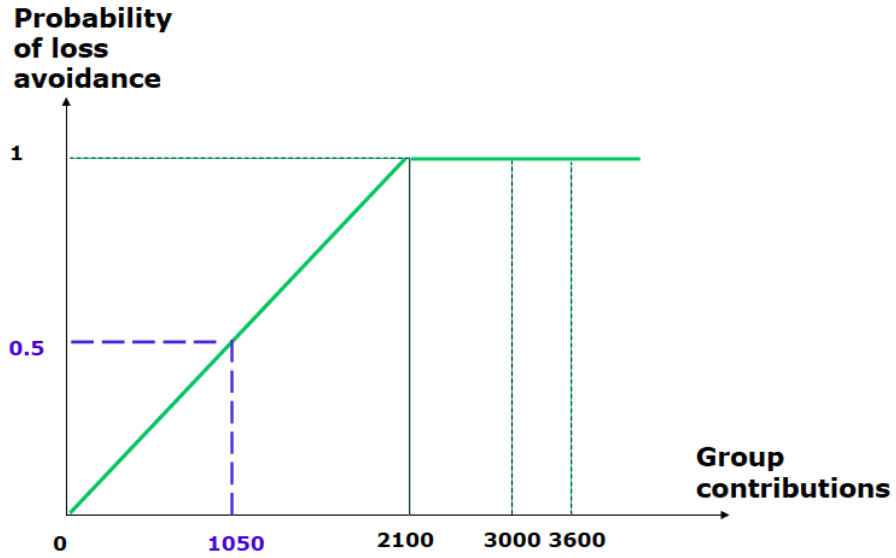
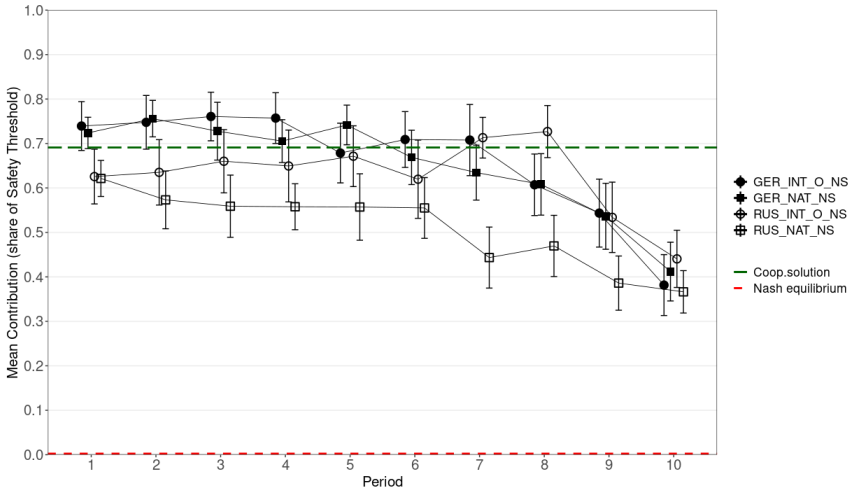


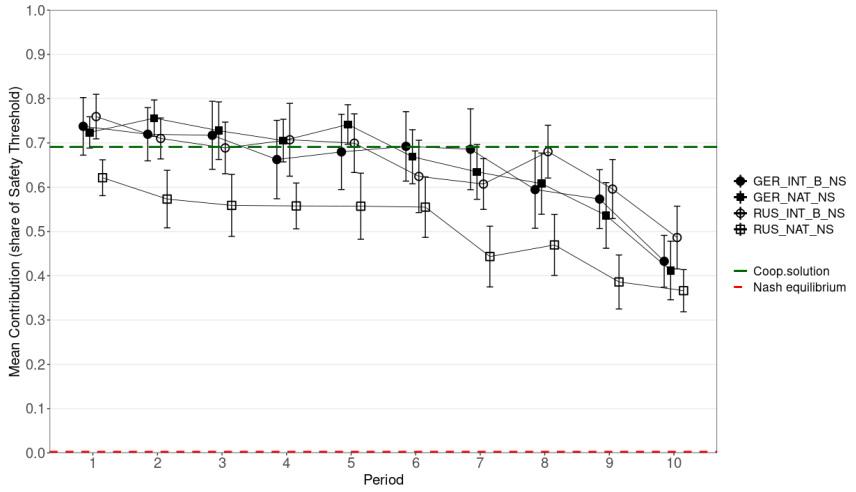
Fig. S4: The loss avoidance scheme. The probability of loss avoidance was proportional to the tokens contributed to the collective fund by the group members. Loss was certain with no contribution, and was avoided with certainty when group contributions equaled the threshold of 2,100 tokens. For instance, if 1,050 tokens were contributed, the probability of loss avoidance would have been 0.5 (see dashed line). The total number of tokens available for contribution by group members was 3,000, while the sum of individual endowments (including tokens available for sanctioning) was 3,600 tokens.

77

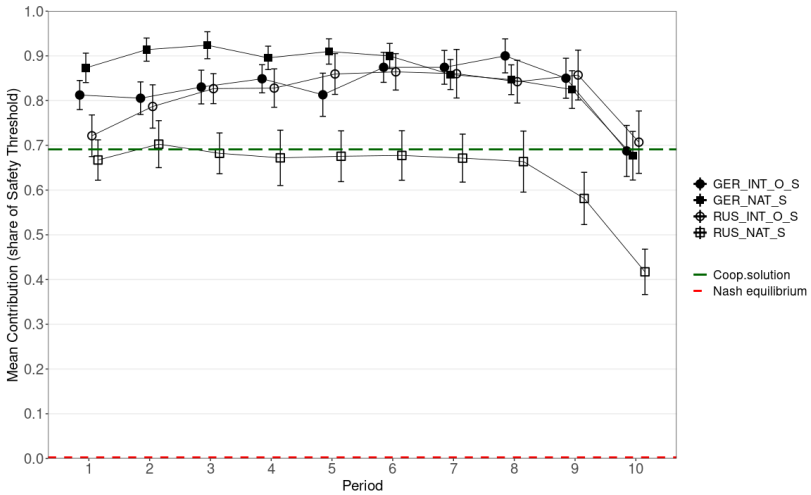
(A): National and International Open treatments without Sanctions



(B) National and International Blind treatments without Sanctions



(C): National and International Open treatments with Sanctions



(D): National and International Blind treatments with Sanctions

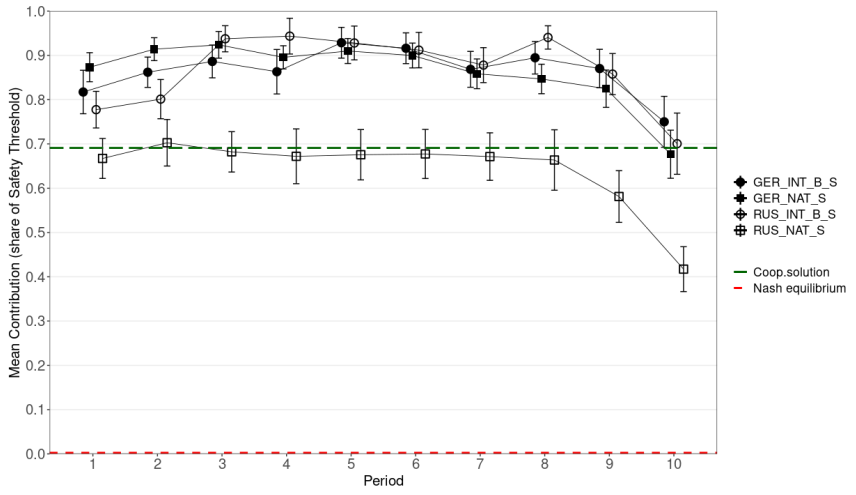


Fig. S5: Evolution of cooperation rate by treatment and nationality. Cooperation rates are expressed as total individual contributions over the 10 rounds as a share of the level associated with the certain safety threshold, which is 350 tokens (assuming that all group-members contribute the same amount). Since the total token endowment available for each individual is 500 tokens, the range of Contribution is [0 , 1.42]. The Cooperative Solution prescribes a contribution level equal to approximately 69% of the certain safety threshold, which is indicated by the green line.

GER_Int_O_NS = German participants' decisions in International Open treatments without sanctions;
GER_Int_O_S = German participants' decisions in International Open treatments with sanctions;
GER_Int_B_NS = German participants' decisions in International Blind treatments without sanctions;
GER_Int_B_S = German participants' decisions in International Blind treatments with sanctions;

RUS_Int_O_NS = Russian participants' decisions in International Open treatments without sanctions;
RUS_Int_O_S = Russian participants' decisions in International Open treatments with sanctions;
RUS_Int_B_NS = Russian participants' decisions in International Blind treatments without sanctions;
RUS_Int_B_S = Russian participants' decisions in International Blind treatments with sanctions;

See Table 1 or Section S7 for definition of other labels.

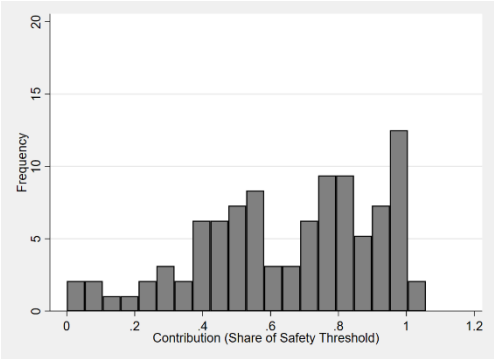
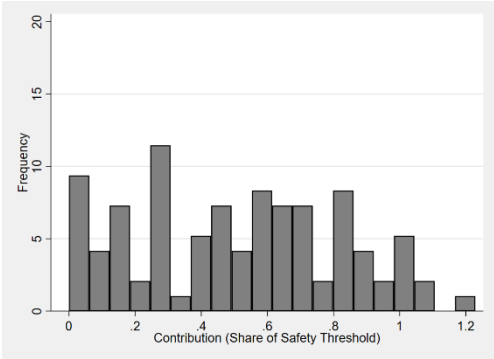
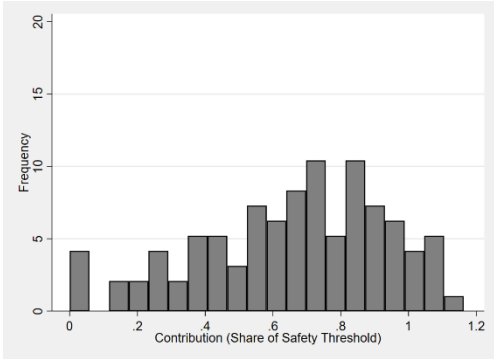
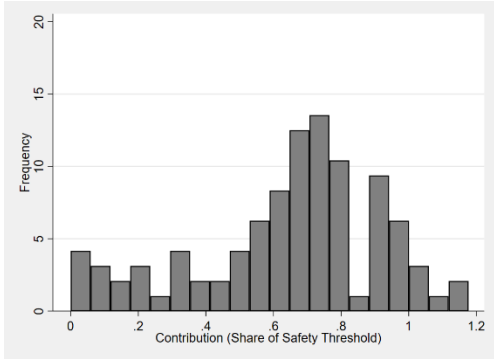
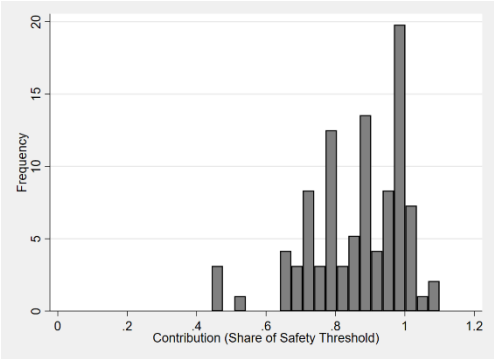
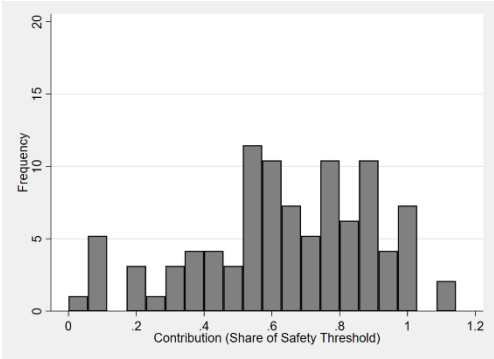
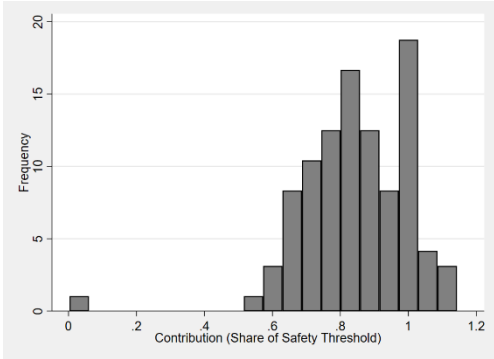
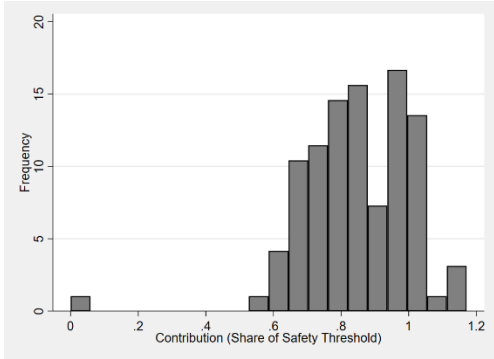
A: National No Sanction - Germany –	B: National No Sanction - Russia	C: International No Sanction – Germans	D: International No Sanction – Russians
			
E: National Sanction - Germany	**F: National Sanction - Russia**	**G: International Germans - Sanction**	**H: International Russians - Sanction**
			

Fig. S6 | Histograms of Total Individual Contributions. Total Individual Contributions over the 10 rounds are expressed in percentages of the level necessary to achieve full risk avoidance (350 tokens), had everyone else contributed the same amount. See footnote to ESM: Fig. S5. For example, if Total Individual Contributions equal 1, it means that a participant contributed 350 tokens, which would produce a PLA=1 had other group members contributed the same amount. Total Individual Contributions are grouped into 20 bins. Frequencies (in percentage terms) are reported on the vertical axis.

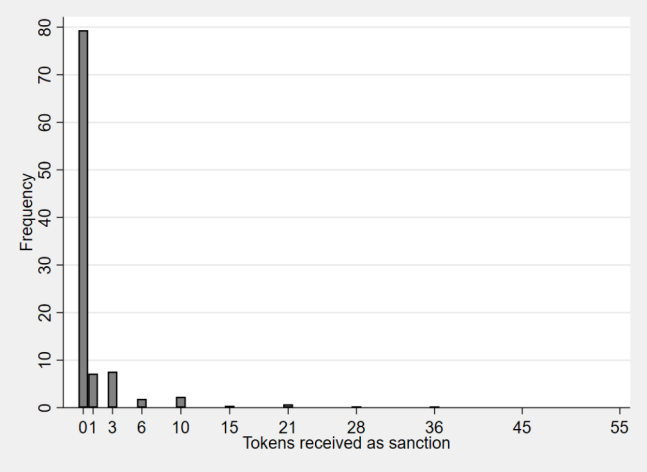
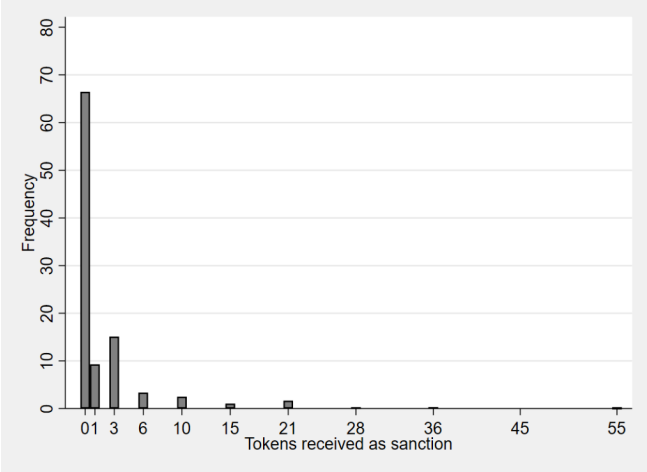
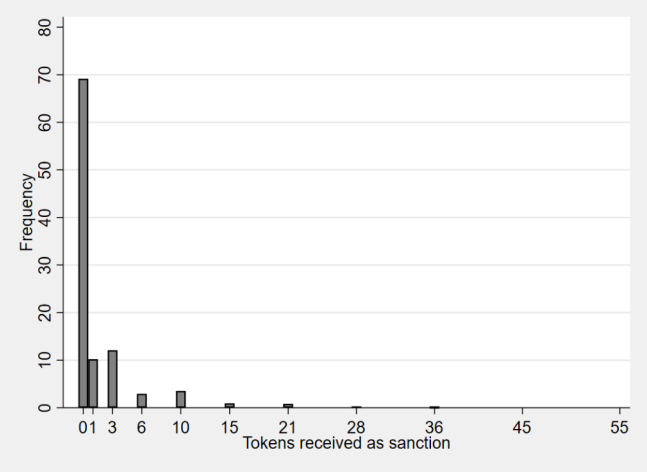
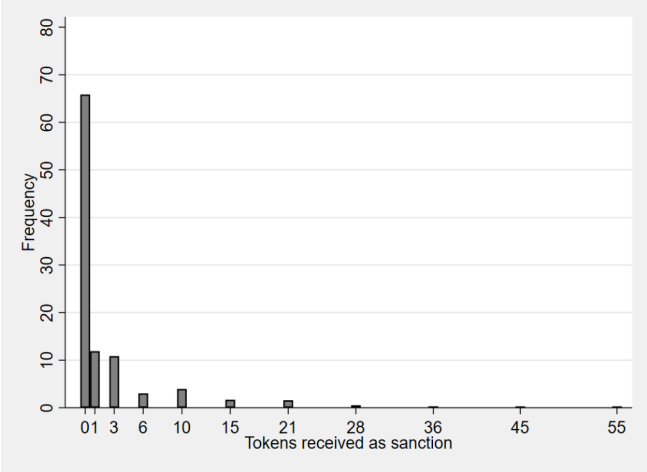
Panel A: National Germany Sanction	Panel B: National Russia Sanction
	
	
Panel C: International Blind Sanction	**Panel D: International Open Sanction**

Fig. S7 | Histograms of tokens lost due to sanctions. The histograms show frequencies of tokens lost to sanctions for each period and individual. Frequencies (in percentage terms) are reported on the vertical axis. See Table S6 for possible levels of sanction losses.

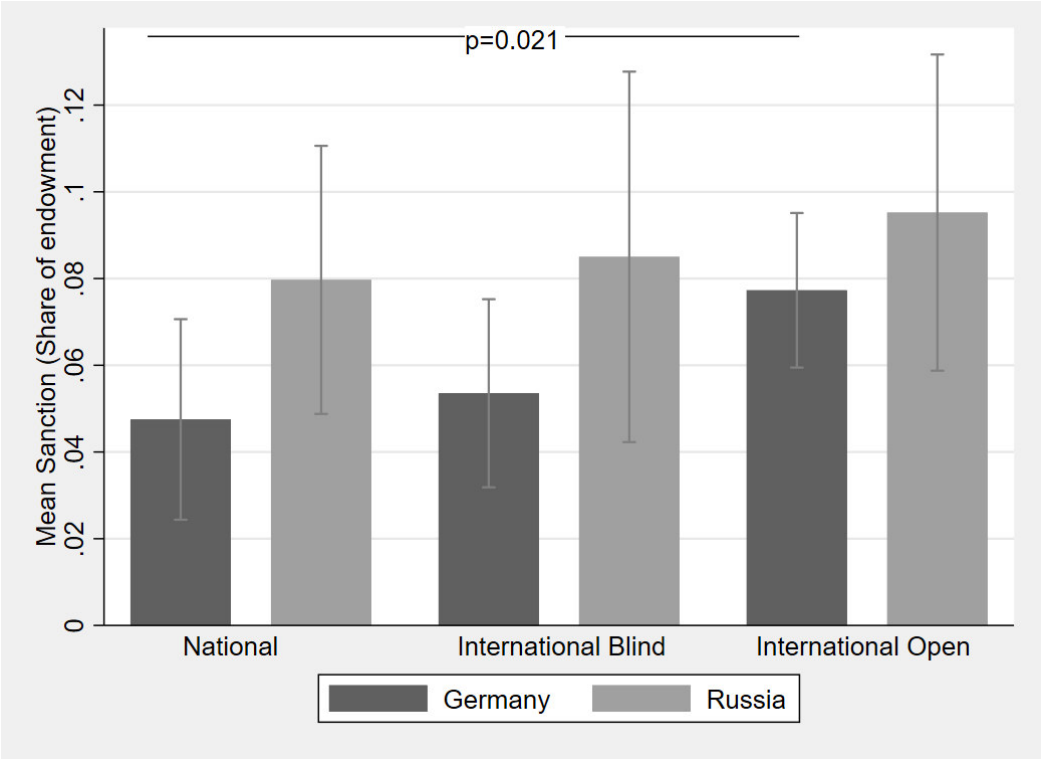


Fig. S8 | Mean number of tokens spent on sanctioning, by treatment and nationality. Means are computed over each (sub)group over the whole 10 rounds, broken down by nationality in international treatments. Error bars are 95% confidence intervals with bootstrapped standard errors (10,000 repetitions).

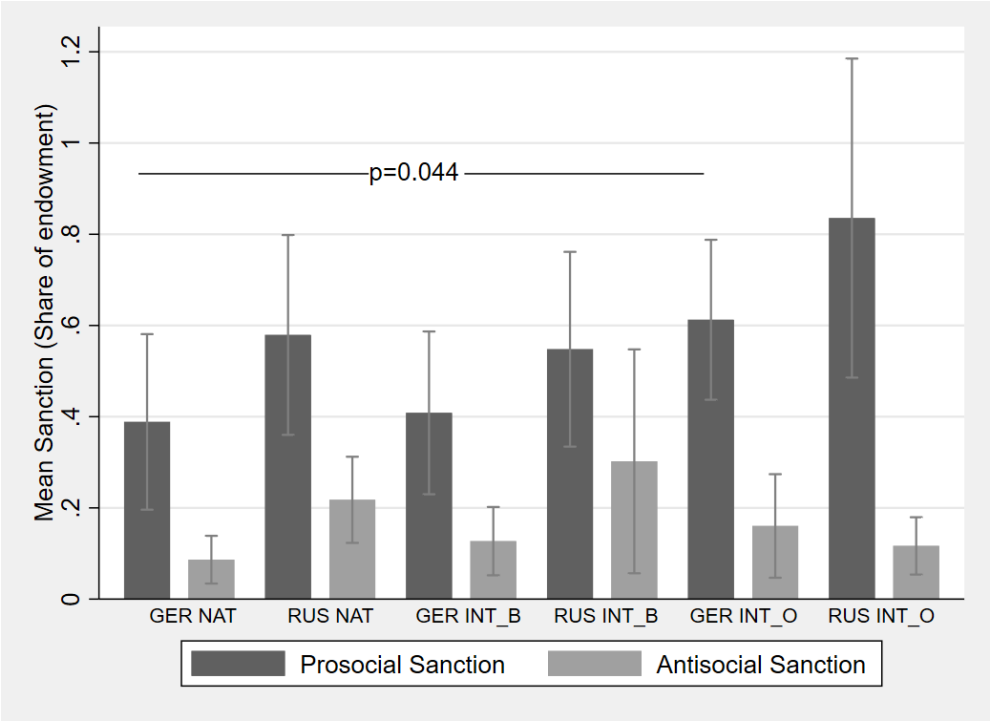
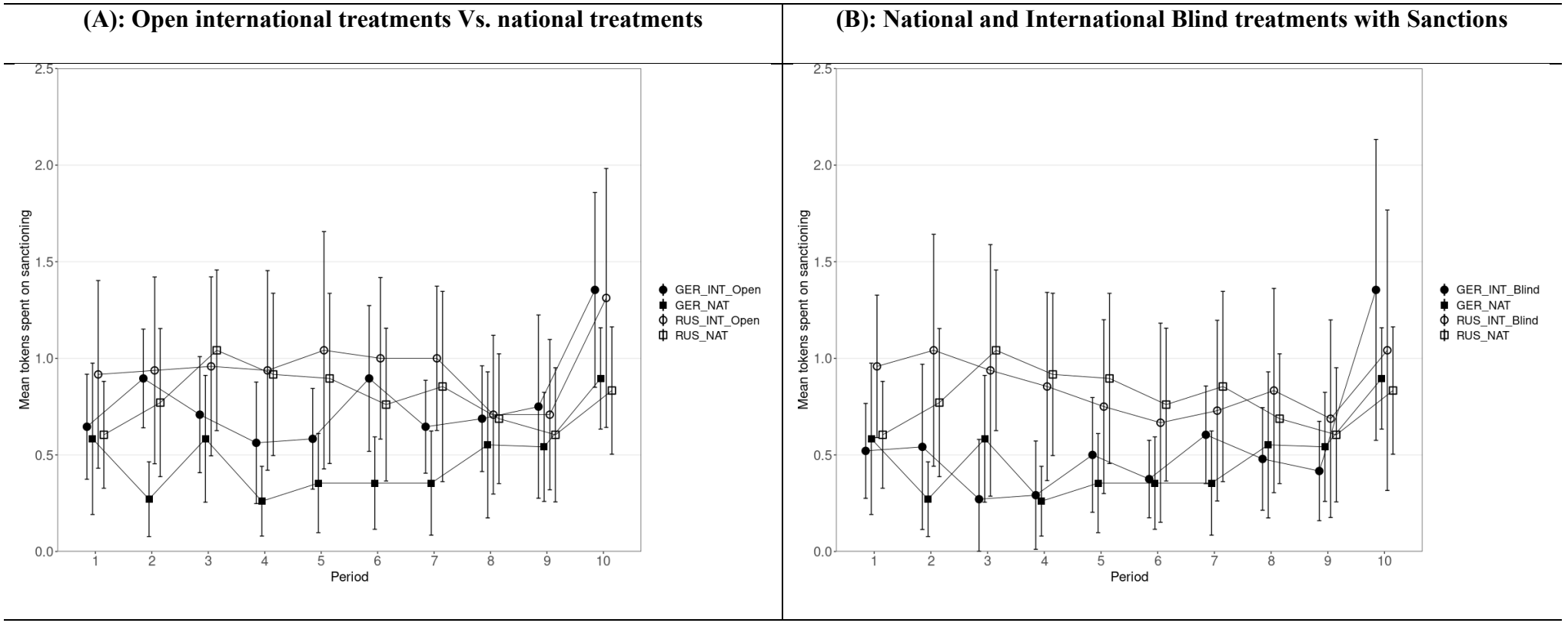


Fig. S9 | Prosocial and antisocial sanctioning (mean number of tokens)**.**
Prosocial and Antisocial sanctioning are defined in Section S1.4.
GER NAT= German National treatment; RUS NAT= Russian National treatment
GER INT_B = German subgroup involved in International Blind treatment.
RUS INT_B = Russian subgroup involved in International Blind treatment.
GER INT_O = German subgroup involved in International Open treatment.
RUS INT_O = Russian subgroup involved in International Open treatment.
Error bars are 95% confidence intervals with bootstrapped standard errors (10,000 repetitions).

109



110 **Fig. S10 | Evolution of sanctioning by treatment and nationality.** See Table 1, Fig. S5, or Section S7 for definition of labels. Error bars are 95% confidence
111 intervals with bootstrapped standard errors (10,000 repetitions).

112

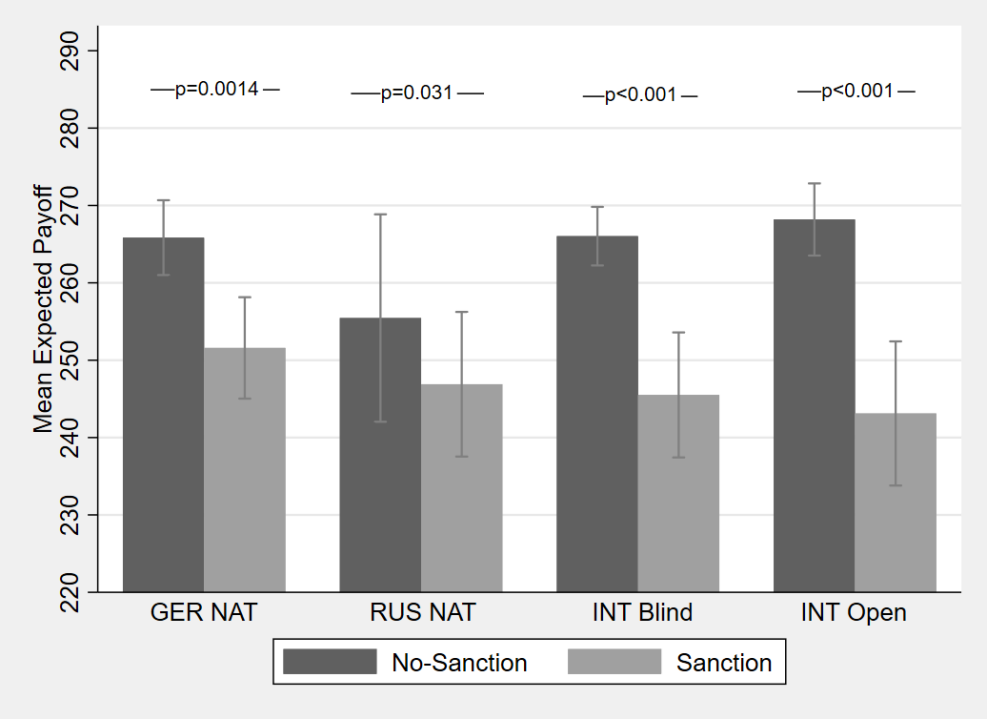


Fig. S11 | Average expected individual payoffs per treatment. See Table 1, Fig. S5, or Section S7 for definition of labels. Only results of pairwise tests between S and NS-treatments for each treatment (e.g. National Germany) are reported. Error bars are 95% confidence intervals with bootstrapped standard errors (10,000 repetitions).

113

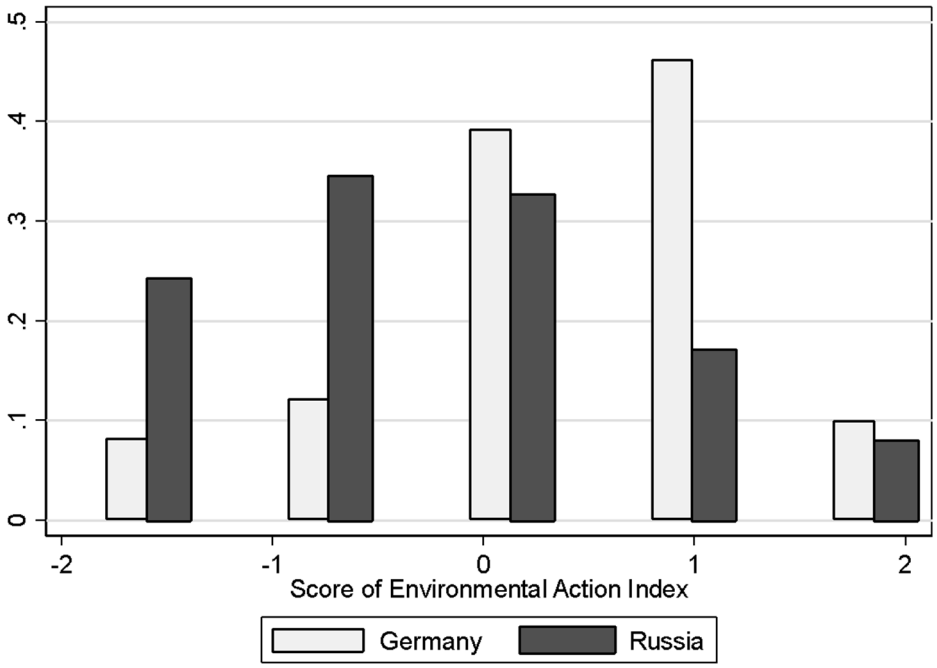


Fig. S12: Histograms of Environmental Action Index. The index is the first principal component of four questions asking whether participants buy environmentally-friendly goods, save water, participate in ecological movements, and are active in recycling (see Section S1.5). The x-axis marks the score of the index. The y-axis is the frequency (in percentage terms) of observations for each possible level of the index.

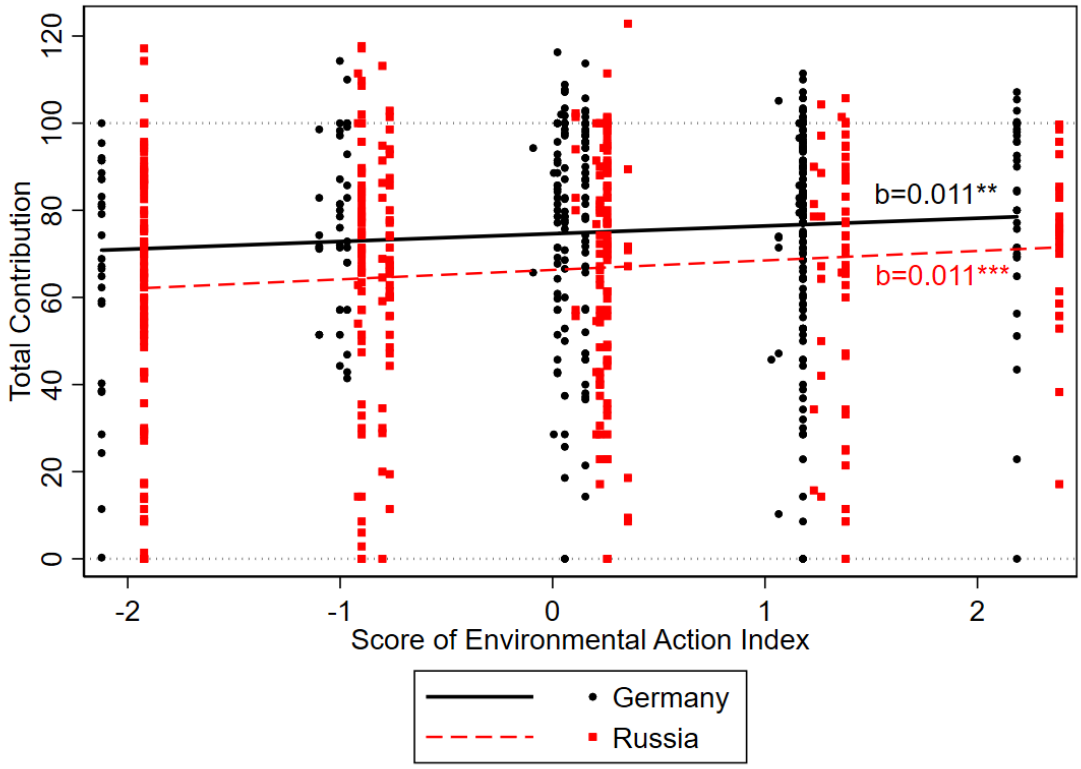


Fig. S13: Relationship between Environmental Action Index and Total Contribution (individual level). Scatterplot of the score of the Environmental Action Index (see Section S1.5 and Fig. S12) on x-axis and of individual Total Contributions over the 10 periods of interaction on y-axis. Environmental Action Index scores for Russia have been shifted rightwards not to have them overlap with those for Germany. Total Contributions are normalised by the level that would have yielded full loss avoidance had anyone in the group contributed that amount (350 tokens). The solid lines are OLS interpolating lines. The coefficient b of the interpolation is reported, together with its significance value. *** $p<0.001$; ** $p<0.01$.

S4 Supplementary Methods

The experiment protocol has been deposited at Protocol Exchange: https://doi.org/10.21203/rs.3.pex-1459/v1 and at Protocolos.io: dx.doi.org/10.17504/protocols.io.bw2ppgdn. The experiment was conducted in z-Tree [65] and the software used in the experiment is available at a repository of the Open Science Foundation, together with the dataset, analyses codes, and experiment protocol and materials (https://osf.io/4s6p3/).

S4.1 Experiments in the social sciences

For several decades the social sciences have applied experimental methods to the study of social interaction in controlled laboratory settings. Our study was characterised by features typical of experiments: (a) Monetary incentivization: Participants received money endowments from the researchers and made decisions on how to allocate these endowments. Participants were then paid the monetary payoffs resulting from their choices and the other group members' choices. Mean earnings were 25.00€ in Germany and 750 Ruble in Russia. (b) Anonymity: Participants' real identity was not revealed to other participants in the course of the experiment. Participants were instead identified through a randomly assigned numeric label so that other participants could reconstruct the "history" of other participants' actions. Reputation-building and revenge were then possible motivations in our experiment. (c) Lack of deception: Participants were never deceived on any aspect of our design. (d) Treatment randomisation: Randomisation of treatments (see Table 1) occurred at the session-level (ESM: section S4.6). Randomisation permits causal inference from the treatments to the main variable of interest – namely, cooperation.

S4.2 Methods to ensure between-country comparability of data

International experimental research is subject to three problems that may

compromise data comparability [66]. We follow relevant literature in responding to such issues [20,34,63,66].

- *Experimenter effects*

It is well-known that personal differences between experimenters conducting research sessions may induce some differences in participants' behaviour. Personal differences include personality or gestural differences, or other physiological differences in, for instance, voice pitch, intonation, and, of course, gender and age, which may ultimately elicit different responses by participants. These effects could not be eliminated, but we strived to minimise them.

Firstly, we produced an experimental script (available at https://osf.io/k4d8w/) that provided a detailed description of the various stages of the experimental session and the instructions to be administered to participants (see the timeline of the experiment in Fig. S14). Each lead researcher (the authors of this paper) read the same instructions from this script, thus ensuring that identical information was given in identical order and in identical format in all the research sessions. Some of these instructions involved PowerPoint© presentations (available at https://osf.io/ch4gd/), which were prepared using the same format for all locations. Since the sessions were run simultaneously, the duration of the various stages of the session had to be approximately the same. Research materials, such as the materials to run the final lottery draw, and the video cameras used for the video links (see next sections) were also the same in all locations.

Secondly, the lead researchers participated in two collective meetings before data collection, in which session procedures were discussed and agreed upon. In one meeting in Moscow, a mock experimental session was conducted by one lead researcher under the observation of all others, in order to make the conduction of the session as uniform as possible.

- *Language effects*

Since a word may have a different nuance, or additional meanings, when translated into another language, language effects may also preempt full comparability of international experimental data. Differences in syntactic rules across languages, and the fact that language expressions ultimately reflect different cultural norms in the way people address each other in different countries, may also introduce some subtle differences in the way people react to the same set of instructions in different languages. In fact, a significant foreign language effect in decision-making has been found [67]. We followed what we believe is the best practice in cross-country and inter-country

experimental research [20,66] and used the back-translation method to make instructions in Russian and German as comparable as possible.

As none of the five authors is bilingual in German and Russian, we elaborated the master version of the instructions in English. Researchers from our team translated this version into their native language. We then asked a professional German-Russian translator to back-translate the Russian version of the instructions into German. This back-translated version was compared with the original German version. Every difference in the two versions was discussed among members of our team and the translator, and the original translations were then adapted to minimise differences in connotation.

- *Currency effects*

Another issue that could hinder comparability is the possibility that the monetary incentives used in different locations were different from each other. We followed standard practice in experimental economics, and formulated instructions referring to 'tokens' rather than to national monetary units. Adjusting the monetary value of a token using the official exchange rate between two currencies is not sufficient, because differences in general price levels between the two countries will alter the purchasing power of a currency when exchanged into another currency. Given that official statistics of Purchasing Power Parity are published with a delay of some years on current prices, we used the standard hourly pay rate for student assistants at universities in each country as the conversion factor to ensure that the monetary value of a token had the same purchasing power in each location. This method is appropriate for university students. This resulted in a token being worth 0.07 Euros in German locations and 2.0 Ruble in Russian locations. In addition, participants received a show-up fee of 5 Euro/150 Ruble.

S4.3 Determination of sample size

We anchored the sample size in our study to the sample size of other studies with a similar design to ours [17,22]. In these studies, the unit of observation is a group of participants, and each group comprises 6-10 participants (we chose the lower bound of 6 for our experiment). These studies had 10 groups per treatment and found a very large effect size for their treatments. In particular, the size of the effect of introducing uncertainty over the safety threshold in one of these studies [22] was Cohen's d=3.59 {m1 = 150.9, m2 = 79.9, sd1 = 7.69, sd2 = 26.90}. We were sceptical that in the context of our study, in which the main treatment concerns the variation in cooperation

in an international environment vis-à-vis a national one, the effect size would have been as large. Therefore, we decided to increase sample size to N=16 per treatment. Ex post power analysis confirmed that our prediction was correct. The sample size requested for Type-1 error = 0.05 and for Power = 0.80 to detect a significant difference in the means observed in one of our key treatments (the difference of cooperation in the International Open treatment and the National Russian treatment under sanctions, where {m1 = 22.4375, m2 = 28.7875, sd1 = 6.9067, sd2 = 4.6133}) is N=15, which is very close to our choice of N=16. The size of this effect is Cohen's d=-1.16.

S4.4 Ethical approval and data protection

Since our research could not provide any harm to participants and did not involve any medical treatment, the approval by an ethics committee or institutional review board was waived by our universities. The experiments were run according to the ethical standards of the experimental economics profession that do not allow deception. We followed standard procedures when dealing with human subjects, and asked every participant to read an information sheet and sign an informed consent form. Data were fully anonymized upon starting the session, as participants were assigned ID codes as soon as they entered the experiment room, and every one of their decisions and answers to the questionnaire was recorded through that number. Payments were paid in cash inserted in a sealed envelope at the end of the session. Participants were asked to sign a receipt, but this was not handled by researchers but was sent to the university administrative office. No participant refused to sign the informed consent form or decided to drop out of the study, even if it was clearly stated that this was possible at any time during the session.

S4.5 Generalisability of results from student samples

Sampling university students is subject to three types of biases: (a) A self-selection bias concerning the university student population; (b) A bias caused by participants performing more socially desirable behaviour when interacting in the lab than in real-life; (c) A lack of representativeness of university student vis-à-vis the general population. Based on specifically designed experimental studies, we conclude that (a) the bias between university students who self-select into experimental studies and the full population of university students appears to be negligible [68]. (b) Even if

some studies show more socially desirable behaviour in experiments than in real life [69-70], other studies do not detect this effect [68]. Most importantly, experiments permit tightly controlled variation in the main parameters of the interaction, thus enabling causal inference. This would be in most cases impossible in natural settings [71]. (c) Even if several studies find that adult samples behave more prosocially than students' sample [72-73], this does not prevent causal inference as long as treatment effects are also distorted by the type of population being sampled. It is then reassuring that correlations across a broad range of variables have similar size in university students' samples and samples representative of the general population [68]. In fact, less noise and fewer cognitive errors have been found in student samples than in representative adult samples [68], which suggest that students' sample may be more reliable than representative samples for hypotheses testing. To further test the generalizability of our results, we conducted an out-of-sample estimation to simulate the size of the treatment effects had we used representative samples of the populations. In ESM: Section S1.9 and Table S15, we have reported the results from an out-of-sample exercise, estimating treatment effects for a hypothetical adult sample upon inference from our student sample. We find that the main treatment effects – in particular, the higher cooperation rate in international treatments with sanctions compared to national treatments – would hold even with nationally representative samples.

S.4.6 Experimental protocol

- *Randomisation*

Randomisation occurred at the session level. Since we wanted to achieve a fully balanced sample across treatments, we did not randomise a treatment for each session, but rather we followed a pre-fixed sequence that alternated treatments. The treatment sequence had to take into account various constraints. One constraint was that our International "Blind" treatments (where participants were not informed that they were interacting with people from another country) had to be conducted before the International "Open" treatments. Had we done differently, "contagion" effects across participants from different sessions may have affected the internal validity of the Blind treatments, because some students may have inferred that other participants were from another country. Our strategy was overall successful because most participants revealed that they expected the other laboratory to be located within their country in the Blind treatments (ESM: Table S4). Other constraints had to do with the university academic

calendars, as students were not present on campus out of term. We balanced the assignment of treatments to starting times, to avoid that, say, all sessions belonging to one treatment were run in the morning, while all sessions relative to another treatment were run in the afternoon. This aspect of the design should prevent that treatment effects were confounded with self-selection into particular times of the day.

- *Recruitment and admission*

32 sessions were conducted between November 2016 and February 2017. We tried to run the sessions in the shortest possible time, compatible with the university calendars. Participants were recruited via email in Tomsk and via the recruiting systems BeLab-System in Moscow and hroot [74] in Kiel and Bonn. Upon arrival, we checked students' passports and admitted only national passport holders to the session. Participants were given an information sheet and were asked to sign an informed consent form before entering the laboratory. Upon arrival, participants were randomly allocated to individual cubicles divided by opaque separators (Fig. S15) to ensure the privacy of decisions. They were randomly divided into groups of six with three group members each being from two different locations in Germany and/or Russia depending on the treatment.

- *Instructions*

All sessions were computerised using the experimental software z-Tree [65] (programs are available at https://osf.io/x82j5/). Participants from the two locations interacted via the Internet and took their decisions at the same time. They received equivalent experimental instructions in their respective native language. Participants were informed that all participants would take their decisions simultaneously and would be provided with equivalent instructions. (See the English translation of the instructions in Section S5).

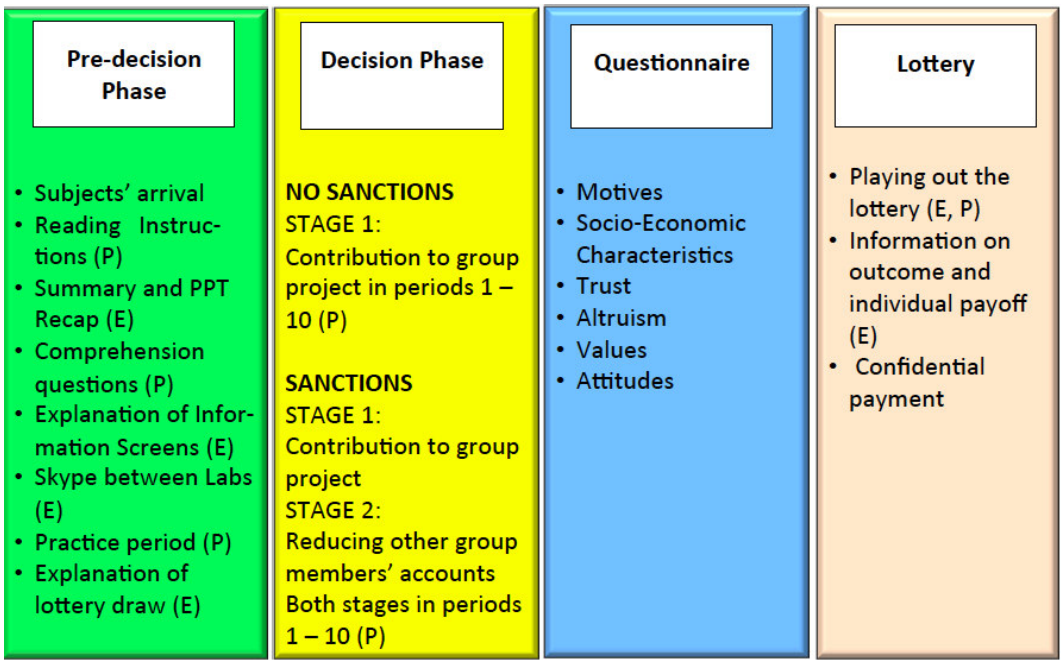


Fig. S14: Timeline of the experiment. (P): Participant's activity/decision; (E): Experimenter's activity

Participants were given ample time to read the instructions and ask clarifying questions which were answered in private. To ensure that participants understood the decision task and the procedure of the experiment, we summarised the instructions in a PowerPoint© presentation with text in German or Russian (available at https://osf.io/ch4gd/). We also made clear that due to our confidential payment method we were not able to trace any individual participant's decisions. Participants then had to answer a set of comprehension questions on their computer screens – showing German text in the German locations and Russian text in the Russian locations also in international treatments. The decision stage did not start unless all participants had answered all control questions correctly.

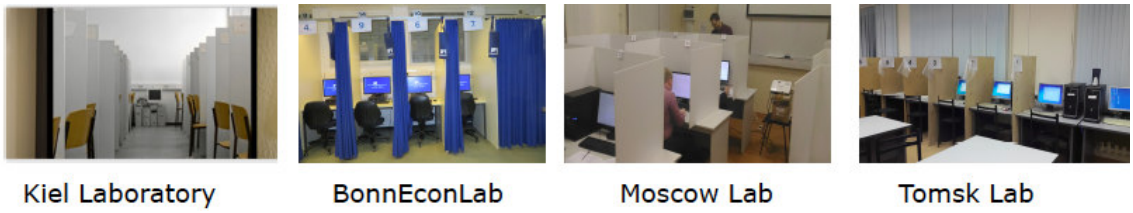


Fig. S15: The four experimental laboratories.

Before entering the decision part of the experiment, participants were presented with an information recap in PowerPoint© to make them familiar with the information options provided throughout the experiment.

- *Video Conference link*

To attenuate possible suspicions on the existence of the other lab, we set up two Skype© connections during the session, lasting a few minutes each. Lead researchers would briefly greet each other and introduce the other participants on a large projector screen visible to all participants. Participants were not allowed to talk or communicate with each other in this phase – as well as in any other phase of the research session. To show that the interaction was occurring in real time, we followed previous research and asked some participants in one location to state some numbers [75]. Such numbers were communicated via the Internet to researchers at the other location, who then wrote these numbers on a slip of paper and showed them on the projector screen through the videoconference link. The same procedure was repeated at the other location. An identical protocol was repeated in all sessions, with the exception that researchers communicated in the respective national languages in the National treatments, interacted in English in the International Open treatments, while the Skype© link was muted in the International Blind treatments. We believed that this set of procedures was best suited to fully assure participants that they were not being deceived and that all the information given in the instruction was truthful.

- Information and feedback

Participants interacted over ten periods with the same partners in real-time via the Internet. Interactions were anonymous, but each group member could be identified by a number ranging from 1 to 6. Since participants knew that group members labelled from 1 to 3 were from one location while those labelled from 4 to 6 were from the other location, participants could infer other group members' location from their numeric label. At the end of each round of contributions, participants received information on each of the other group-members' contribution in the previous round and over all the previous rounds, as well as the current probability of loss avoidance determined by total contributions. In S-treatments, participants also received information on the sanctions assigned by a group member to any other group member.

- *Decisions*

After the video conference link, participants completed a practice period on their computers. In order not to bias actual experimental decisions, participants were not communicated others' decisions in the practice period but rather were only allowed to get familiar with the commands of the software. Afterward, the experimenters in both locations explained how the lottery would be implemented. Finally, participants made their decisions in periods 1 to 10 in the NSNS)- or S-treatments.

To illustrate the participants' decision task, Fig. S16 provides the decision screen for the contribution decision in Stage 1. In addition to making their decisions, each member was informed about the contributions of all six group members as well as about the tokens in each of their personal accounts, both accumulated over the previous periods. Furthermore, they saw the total number of tokens contributed to the project and the current probability that the loss event will not occur.

After participants had taken their decisions they could get visual information on each group member's contributions in each of the previous periods (Fig. S17). In NS-treatments, the period ended at this point and each participant was informed about their contribution in the current period as well as about everyone's personal account in tokens at the end of the previous and the current periods.

In S-treatments, participants entered Stage 2 and made their decision on how many tokens they wanted to spend to sanction each of the other group members. Before having done so they could retrieve information on each group member's contributions in each of the previous periods (Fig. S17) and in the current period (Fig. S18), the accumulated number of tokens in each group member's personal account, and the number of tokens each group member spent in the last period on each of the other group members to reduce that person's personal account.

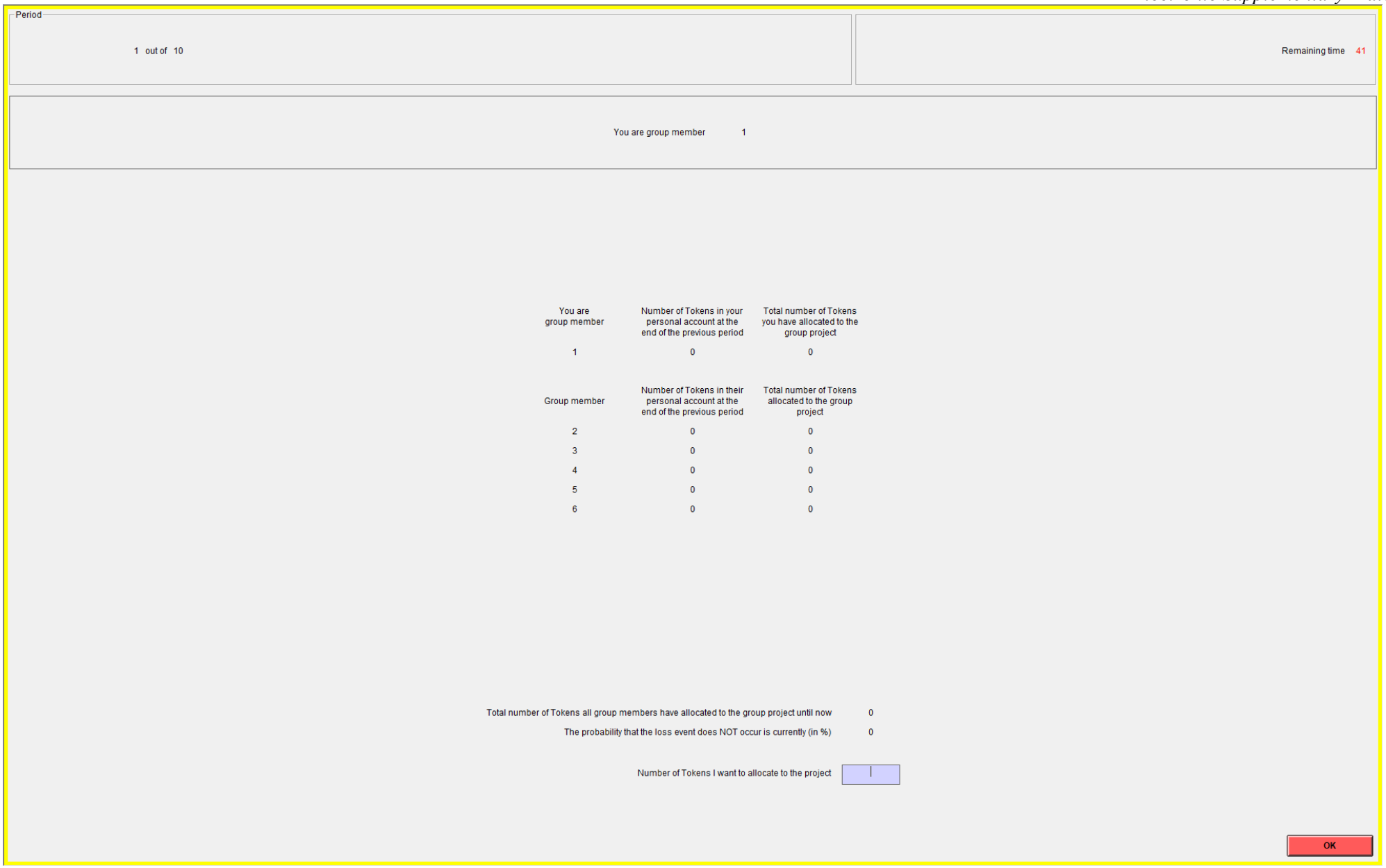


Fig. S16 | Participants' decision screen and information in Stage 1.

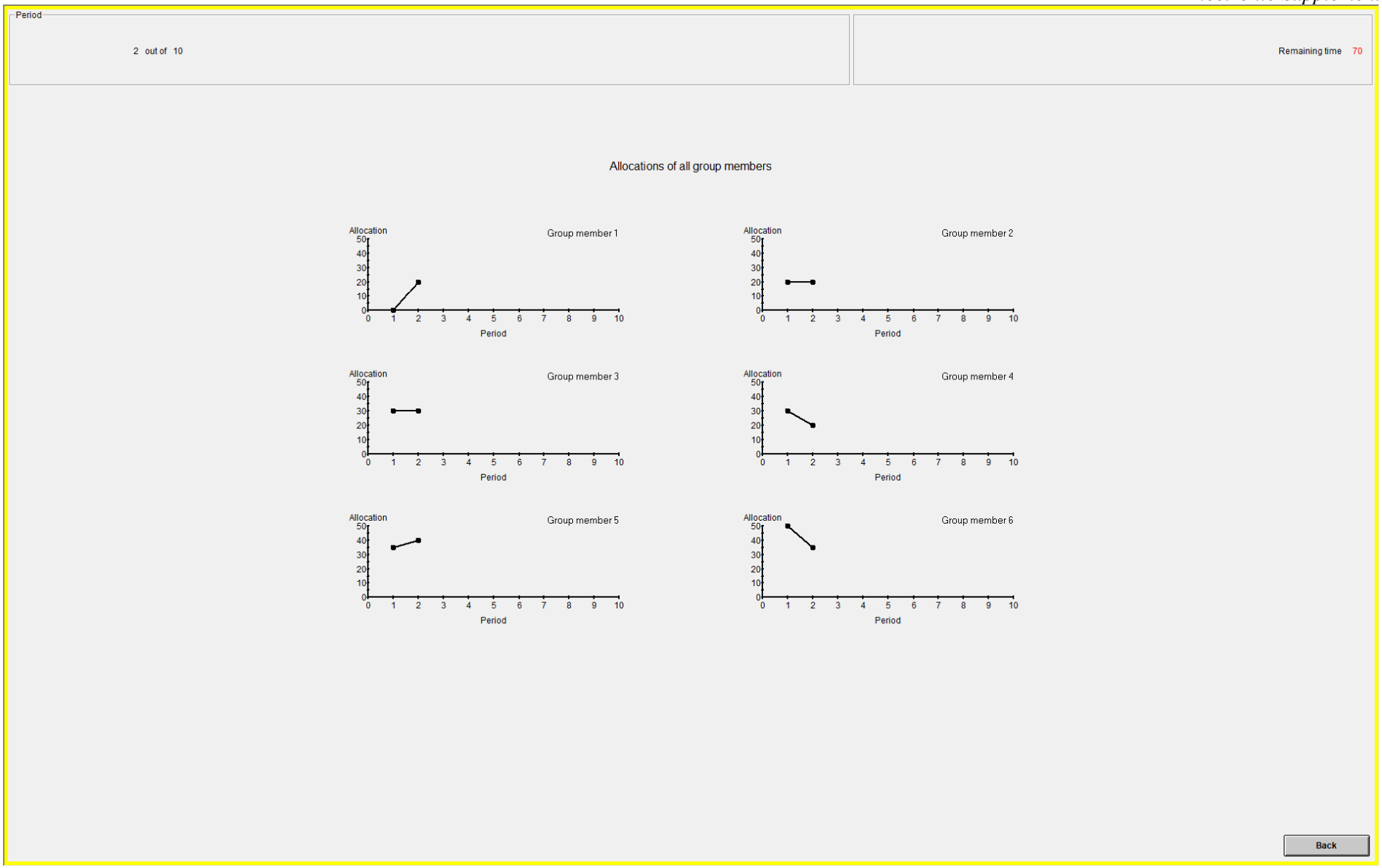


Fig. S17: **Graphical information on each group member's contributions in each of the previous periods.**

Fig. S18: Participants' decision screen and information in Stage 2.

At the end of Stage 2 of each period, participants received the same information as in NS-treatments and learned the number of tokens spent on others and deducted from their own account.

- *Final procedures*

Having finished the experimental tasks, the lottery to decide whether the loss event would occur was played out. From a bag containing lottery chips numbered 1 - 100, one chip was drawn by a participant. If the number was larger than the percentage *x* of the target amount the group members had contributed to the group project, the loss event occurred and 75% of the amount collected in each group member's personal account was lost. If the number drawn was smaller or equal to *x,* the loss event did not occur and each group member was paid out the total amount in his or her personal account. This procedure was repeated for each of the four groups participating in a session. The outcomes of the lottery draws were transmitted via Skype© to both participating labs but information on the lottery outcome relevant for a specific group was not made available to the participants until they had filled in a non-incentivized questionnaire on social characteristics, risk attitudes [55], personal values [76] and other questions taken from the World Value Survey (see Section S6). The survey questions were available and externally validated in both languages.

Finally, we applied an anonymized payment procedure by distributing the payments from the experiment plus the show-up fee and receipts in an envelope marked with the cubicle number. Participants took the money, signed the receipt, confidentially put the receipt into a box, and left the laboratory. All features of the experimental design and procedure were common knowledge and did not raise any questions. Sessions lasted about 2 hours on average. Mean earnings were 25.00€ in Germany and 750 Ruble in Russia (12.5€ at the time of running the experiment) including the show-up fee (see Section S4.1).

S5 Instructions

Note: No-Sanction treatments ended after Stage 1.
[Sentences in brackets: Change in instructions according to specific treatments.]

General instructions to the participants

Welcome to this experimental session. You will take a sequence of decisions and you have the opportunity to earn money. How much money you earn will depend both on your decisions and the decisions of other participants. It is therefore very important that you read these instructions with care.

Your total payoff will be paid in cash at the end of the experimental session.

Both your decisions and your payoffs are anonymous, that is, no other participant will be able to associate this information with a specific person during or after the experimental session. We commit to treat your decisions confidentially and analyse them anonymously.

These instructions are for your private use only. During the whole experimental session, it is not allowed to communicate with other participants. If you violate this rule, you may be dismissed from the experiment and forfeit all payments.

Should you have any questions, please raise your hand. We will then come to your workstation and answer your questions in private.

During the session we will not talk in terms of [Euro/Ruble], but in terms of Tokens. That means, your entire payoff will first be calculated in Tokens. At the end of the experimental session your total amount of Tokens will be converted to [Euro/Ruble] at the following rate:

1 Token = 0.07 Euro/2 Ruble .

In addition you will receive a show-up fee of **5** Euro/ **150** Ruble.

At the end of the session each participant will be paid in private.

The participants of this experimental session are randomly divided into groups of six. You will therefore be in a group with five other participants. You and two other participants are together in this room and the other three participants are students

- [National treatment: from another German city. / another Russian city.]
- [International Blind treatments: (in both locations): another city[6]].
- [International Open treatments: (in German location): a Russian city. / (in Russian location): a German city].

[**International open only**: Note that for participants in Germany 1 Token = **0.07** Euro. For Russian participants 1 Token = **2.0** Ruble. The exchange rate is such that equal amounts of Tokens have equivalent purchasing power in both countries.]

Participants in both universities interact via the internet. All participants take their decisions at the same time and are provided with equivalent instructions**.**

We will set up a Skype connection later to show you that you are interacting in real time with participants from another city.

Your decision tasks in this experiment

There are 10 periods in this session and the composition of the groups will stay the same for all periods. Each group member is identified by a specific number (1, 2, 3, 4, 5 or 6). The identification number for each group member stays the same in all periods.

In each of the 10 periods in this session you will receive an initial sum of 60 Tokens.

[6] Three German participants were in fact matched with three Russian participants, but this was not revealed. See Table 1, Section S4.4, and Table S4.

[Sanction treatments: Each period is divided into a first stage (Stage 1) and a second stage (Stage 2). In each period, you can use 50 Tokens to make decisions in Stage 1, and 10 Tokens to make decisions in Stage 2.

We now explain your task in Stage 1. Later we will explain your task in Stage 2].

YOUR TASK [Sanction Treatments: IN STAGE 1]

You have to decide how many of the 50 Tokens you want to allocate to a **group project** (see below) and how many you want to keep for yourself. What you keep for yourself will be collected in your **personal account**, and shall immediately be added to it. As you decide how many Tokens to allocate to the project, you also decide how many Tokens you keep for yourself. This is:

Amount added to your personal account =
50 Tokens minus the amount you allocated to the group project.

You can only choose integer numbers.

No one of the group members will observe others' decisions when making their own.

All the Tokens allocated to the project by the six group members during the 10 periods will be accumulated. If the group members altogether allocate **at least 2,100 Tokens to the project by the end of the last period**, each group member will be paid out what he or she has collected in his/her personal account over the 10 periods (plus the show-up fee).

If the group members altogether allocate less than 2,100 Tokens to the project, then **a loss event may occur with some probability**. If the loss event occurs, **75% of the total amount each group member has collected in his/her personal account over the 10 periods will be lost**. The remaining 25% will be paid out to each group member (plus the show-up fee).

The probability with which the loss event does not occur depends on the **percentage of the target amount of 2,100 Tokens that the group allocates to the project**. The more tokens **the group** allocates to the project **in total**, the higher the probability that the loss event will not occur.

In general terms, if the group members allocate in total x% of the target amount of 2,100 Tokens to the project by the end of period 10, the loss event will not occur with a probability of x%, and will occur with a probability (100-x)%.

For instance, if the group members allocate a total of 1,050 Tokens to the project, the loss event will not occur with a probability of **50%** – because 1.050 is 50% of 2100 Tokens:

$$Probability\ that\ the\ loss\ event\ does\ not\ occur = \frac{1050}{2100} = 0.5(50\%).$$

If the group members allocate in total 1,890Tokens to the project, the loss event will not occur with a probability of **90%** (because 1,890 Tokens is 90% of 2,100 tokens) and will occur with the residual probability of 10%.

$$Probability\ that\ the\ loss\ event\ does\ not\ occur = \frac{1890}{2100} = 0.9(90\%)$$

If the group members allocate in total 0 Tokens to the project, the loss event will occur for sure.

If the group members allocate in total 2,100 Tokens, **or more**, to the project, the loss event will **not** occur for sure.

Note that **the Tokens allocated to the project will never be returned to anyone, regardless of whether the loss-event occurs or not**. In particular, if the group allocates more than 2,100 Tokens to the project, the Tokens in excess of 2,000 Tokens are also not going to be returned to anyone. For instance, *if 2,200 Tokens are allocated to the project, no one receives back the 2,200 Tokens.*

YOUR TASK IN STAGE 2

In each of the 10 periods, you have to decide how many of the **10** Tokens you receive in Stage 2 you want to **spend to reduce the number of Tokens in other group members' personal accounts** or how many you want to **put in your personal account**. Any Token(s) you put in your personal account will be immediately added to it. Any Token(s) you spend will reduce the personal account of some other group member(s). By spending their Tokens other group members can also reduce your personal account. Or they can leave it unchanged. They can also reduce others' personal accounts or leave them unchanged.

How does this work? You can spend 1 or 2 Tokens to reduce the personal account of any other of your group members, or you can decide to spend nothing. How many Tokens will be deducted from the other members' personal accounts depends on how many group members decide to spend their Tokens, according to the following table:

Total number of Tokens spent to reduce one group member's personal account by the other five group members	Number of Tokens deducted from this group member's personal account
0	0
1	1
2	3
3	6
4	10
5	15
6	21
7	28
8	36
9	45
10	55

You will notice that the number of Tokens deducted from a group member's personal account will increase over-proportionately if other group members spend more Tokens to reduce that group member's account.

If you and any other group member do not spend any Token(s), no Tokens will be deducted from any other group member's personal account. If you spend 1 Token to reduce the personal account of a given group member, and nobody else spends any token, then this group member's personal account will be reduced by 1 Token.

If you spend 2 Tokens to reduce the personal account of a given group member, and nobody else spends any token, then this group member's personal account will be reduced by 3 Tokens.

Likewise, if you spend 1 Token to reduce the personal account of a given group member, and another group member spends 1 Token, and nobody else spends any token, then this group member's personal account will also be reduced by 3 Tokens.

If other group members spend a total of 3 Tokens to reduce your account, this will decrease by 6 Tokens. If other group members spend 5 Tokens your account will be reduced by 15 Tokens.

Note that the amount of Tokens in your personal account cannot ever become negative. If the total number of Tokens that you spend and others want to reduce from your personal account exceed what you actually have in your personal account, your personal account will go to zero, but will not become negative.

At the end of Stage 2 of each period, the total amount of Tokens in your personal account

=

Tokens collected in your personal account by the end of the previous period

(This is 0 in the first period)

+

60 Tokens you have received at the beginning of this period

-

Tokens you have allocated to the project in Stage 1 of this period

-

Tokens you have spent in Stage 2 of this period

-

Tokens deducted from your personal account in Stage 2

OR **ZERO TOKENS**, IF THE SUM OF ALL TERMS ABOVE IS NEGATIVE.

Before making your decisions, you will receive information on others' decisions. We will explain to you this information later in detail.

How it is determined whether the loss event occurs will be explained later.

When you are finished reading these instructions, please click the OK button.

Comprehension Questions

[No-Sanction Treatments:]

Part 1

1. If you or another group member contributes more Tokens to the project does the probability that the loss event does not occur rise, decrease or stay the same?
   a. The probability rises.
   b. The probability decreases.
   c. The probability stays the same.

2a. Suppose that over the 10 periods group member 1 has contributed a total of 350 Tokens and group member 5 has contributed a total of 150 Tokens to the project. And suppose the loss event does not occur. Which of the two group members will finally receive a higher payoff?

   i. Group member 1 receives a higher payoff.
   ii. Group member 5 receives a higher payoff.
   iii. Both group members receive the same payoff.

2b. Let us now assume that the loss event does occur. Which of the two group members will finally receive a higher payoff?

i. Group member 1 receives a higher payoff.
ii. Group member 5 receives a higher payoff.
iii. Both group members receive the same payoff.

3. Tokens that are contributed to the project will at the end of the Session
   a. ...not be paid back to those group members who had contributed them.
   b. ... be paid back to those group members who had contributed them.
   c. ... only be paid back if the loss event does not occur.

Part 2

4. Suppose that over the 10 periods the following amounts have been contributed to the project in total:
   - Group member 1 has contributed 500 Tokens,
   - Group members 2, 3 and 4 each have contributed 100 Tokens,
   - Group member 5 has contributed 250 Tokens,
   - Group member 6 has contributed 0 Tokens.

   a. What is the probability that the loss event does not occur?
   b. Assume that the loss event does not occur. What is group member 1's final payoff in Tokens?
   c. What is group member 2's final payoff in Tokens?
   d. What is group member 6's final payoff in Tokens?
   e. Assume now that the loss event does occur. What is group member 1's final payoff in Tokens?
   f. What is group member 2's final payoff in Tokens?
   g. What is group member 6's final payoff in Tokens?

Part 3

5. Suppose the amounts are like in the previous example, yet group member 1 contributes nothing instead of 500 Tokens as before. Therefore, in total the following amounts have been contributed to the project:
   - Group member 1 has contributed 0 Tokens,
   - Group members 2, 3 and 4 each have contributed 100 Tokens,
   - Group member 5 has contributed 250 Tokens,
   - Group member 6 has contributed 0 Tokens.

a. What is the probability that the loss event does not occur?

b. What is group member 1's final payoff in Tokens if the loss event does not occur?

c. What is group member 1's final payoff in Tokens if the loss event does occur?

Sanction Treatments

Part 1

1. If you or another group member contributes more Tokens to the project does the probability that the loss event does not occur rise, decrease or stay the same?

   a. The probability rises.

   b. The probability decreases.

   c. The probability stays the same.

2.a. Suppose that over the 10 periods group member 1 has contributed a total of 350 Tokens and group member 5 has contributed a total of 150 Tokens to the project. And suppose further that no group member has spent any Tokens on reducing the number of Tokens in any other group member's personal account. Thus, all group members keep the 100 Tokens from Stage 2. Moreover, assume that the loss event does not occur. Which group member will finally receive a higher payoff?

   i. Group member 1 receives a higher payoff.

   ii. Group member 5 receives a higher payoff.

   iii. Both group members receive the same payoff.

2b. Let us now assume that the loss event does occur. Which group member will finally receive a higher payoff?

   i. Group member 1 receives a higher payoff.

   ii. Group member 5 receives a higher payoff.

   iii. Both group members receive the same payoff.

3. Tokens that are contributed to the project will at the end of the Session

   a. ... not be paid back to those group members who had contributed them.

   b. ... be paid back to those group members who had contributed them.

   c. ... only be paid back if the loss event does not occur.

4. Suppose that in a given period:

   - Group member 2 spent 2 Tokens,

   - Group member 3 spent 2 Tokens,

- Group member 4 spent 1 Token,
- Group member 5 spent 1 Tokens,
- Group member 6 spent 0 Tokens,

on reducing the number of Tokens in the personal account of group member 1.

- By how many Tokens is the personal account of group member 1 reduced due to other group members spending Tokens on reducing the personal account of group member 1? (Note: Use the table on page 4 of the Instructions).
- By how many Tokens is the personal account of group member 2 reduced in the given period?

Part 2

5. Suppose that over the 10 periods the following amounts have been contributed to the project in total:
   - Group member 1 has contributed 500 Tokens,
   - Group members 2, 3 and 4 each have contributed 100 Tokens,
   - Group member 5 has contributed 250 Tokens,
   - Group member 6 has contributed 0 Tokens.

   a. What is the probability that the loss event does not occur?

   b. Suppose that no group member has spent any Tokens on reducing the number of Tokens in other group members' personal accounts. Thus, all group members' personal accounts will be increased by 100 Tokens. Assume that the loss event does not occur. What is group member 1's final payoff in Tokens?

   c. What is group member 2's final payoff in Tokens?

   d. What is group member 6's final payoff in Tokens?

   e. Assume now that the loss event does occur. What is group member 1's final payoff in Tokens?

   f. What is group member 2's final payoff in Tokens?

   g. What is group member 6's final payoff in Tokens?

Part 3

6. Suppose the amounts are like in the previous example, yet group member 1 contributes nothing instead of 500 Tokens as before. Therefore, in total the following amounts have been contributed to the project:
   - Group member 1 has contributed 0 Tokens,

- Group members 2, 3 and 4 each have contributed 100 Tokens,
- Group member 5 has contributed 250 Tokens,
- Group member 6 has contributed 0 Tokens.

a. What is the probability that the loss event does not occur?

b. Suppose that no group member has spent any Tokens on reducing the number of Tokens in other group members' personal accounts. Thus, all group members' personal accounts will be increased by 100 Tokens. What is group member 1's final payoff in Tokens if the loss event does not occur?

c. What is group member 1's final payoff in Tokens if the loss event does not occur?

S6 Questionnaire

N	Question	Answers
1	Age	Open question
2	Sex	O Female O Male
3	Which degree are you attending?	O economics or business O mathematics or engineering O natural sciences O medicine O social sciences O humanities O arts O other; specify
4	Please indicate your grade point average	Open question
5	In which city were you born?	Open question
6	Did you take part in university exchange programs?	Open question
7	If yes, for how long?	Open question
8	How many years overall have you resided outside Russia (for Russian version) or Germany (for German version)?	Open question
9	In which country was your father born?	Open question
10	In which country was your mother born?	Open question
11	Please indicate how many older siblings you have	Open question
12	Please indicate how many younger siblings you have	Open question
13	Are you married?	O yes O no
14	How tall are you?	Open question
15	How much do you know about global warming or climate change?	O A great deal O A fair amount O Only a little

		O Not at all
16	How much do you worry about global warming or climate change?	O A great deal O A fair amount O Only a little O Not at all
17	Do you think that global warming will pose a serious threat to you or your family in your lifetime?	O yes O no
18	Temperature rise is a part of global warming or climate change. Do you think rising temperatures are a result of human activities, a result of natural causes, or both?	O Result of human activities O Result of natural causes O Both
19	Have you avoided using certain products that harm the environment in the past year?	O yes O no
20	Have you been active in a group or organisation that works to protect the environment in the past year?	O yes O no
21	Have you tried to use less water in your household in the past year?	O yes O no
22	Have you voluntarily recycled newspapers, glass, aluminium, motor oil, or other items in the past year?	O yes O no
23	Please answer for each of the following actions whether you think it can always be justified, never be justified, or something in between: O Claiming government benefits to which you are not entitled O Avoiding a fare on public transport O Cheating on taxes if you have a chance O Someone accepting a bribe in the course of their duties O Homosexuality O Prostitution O Abortion O Divorce	Use the following scale, where 1 means “Never justifiable” and 5 means “Always justifiable”

	O Euthanasia—ending the life of the incurable sick O Suicide O For a man to beat his wife	
24	How proud are you to be Russian / German?	O Very proud O Quite proud O Not very proud O Not at all proud
25	People have different views about themselves and how they relate to the world. Please indicate how strongly you agree or disagree with each of the following statements about how you see yourself? O I see myself as a world citizen. O I see myself as part of my local community. O I see myself as part of the Russian (for Russian version) / German (for German version) nation. O I see myself as part of the Commonwealth of Independent States (for Russian version) / European Union (for German version). O I see myself as an autonomous individual.	Use the following scale, where 1 means "Strongly agree", 2 - "Agree", 3 - "Disagree" and 4 means "Strongly disagree"
26	What language do you normally speak at home?	O Russian (for Russian version) / German (for German version) O Other
27	Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?	Use the following scale:10 means that most people can be trusted and 1 means that you need to be very careful.
28	All things considered, how satisfied are you with your life as a whole these days?	Use the following scale: 10 means 'Completely satisfied' and 1 means 'Not at all satisfied'.
29	For each of the following organisations, state how much trust do you have in them. Answer using the following scale, where 1 means "No trust at all" and 5 means "Full trust".	O Armed forces O Police O Press O Television

		O Environmental organisations O Chancellor /President O Parliament O Government O Political parties O Justice system O The Churches O Migrants from other countries O European Union O Russians /Germans O United Nations
30	How do you see yourself? Are you generally a person who is fully willing to take risks or do you try to avoid taking risks?	Please tick a box on the scale below, where 0 means "fully try to avoid risk" and 10 means "fully prepared to take risks"
31	People can behave differently in different situations. How would you rate your willingness to take risks in the following areas? How are you prepared to take risks… O while driving? O in financial matters? O during leisure and sport? O in your occupation? O with your health? O your faith in other people?	Please tick a box on the scale below, where 0 means "risk averse" and 10 means "fully prepared to take risks"
32	How many inhabitants has the town where you lived at the age of 16?	Open question
33	What are your religious views?	O Atheist/agnostic O Catholic O Protestant O Orthodox O Muslim

		O Jewish O Hinduist O Buddhist O Other:
34	Now I will briefly describe some people. Please read each description carefully and tick the box showing how much each person is or is not like you. 1 Thinking up new ideas and being creative is important to him/her. He/she likes to do things in his/her own original way. 2 It is important to him/her to be rich. He/she wants to have a lot of money and expensive things. 3 He/she thinks it is important that every person in the world should be treated equally. He/she believes everyone should have equal opportunities in life. 4 It's important for him/her to show his/her abilities. He/she wants people to admire what he/she does. 5 It is important to him/her to live in a secure environment. He/she avoids anything that might endanger his/her safety. 6 He/she likes surprises and is always looking for new things to do. He/she thinks it is important to do lots of different things in life. 7 He/she believes that people should do what they're told. He/she thinks people should follow rules at all times, even when no one is watching. 8 It is important to him/her to listen to people who are different from him/her. Even when he/she disagrees with them, he/she still wants to understand them. 9 It is important to him/her to be humble and modest. He/she tries not to draw attention to himself/herself. 10 Having a good time is important to him/her. He/she likes to "spoil" himself/herself.	Use the following scale, where -1 means "Not at all similar to me", 0 - "Not similar to me ", 1 - "Somewhat similar to me", 2 - "To an extent similar to me", 3 - "Similar to me", and 4 means "Fully similar to me"

	11 It is important to him/her to make his/her own decisions about what he/she does. He/she likes to be free and not depend on others. 12 It's very important to him/her to help the people around him/her. He/she wants to care for their well-being. 13 Being very successful is important to him/her. He/she hopes people will recognise his/her achievements. 14 It is important to him/her that the government ensures his/her safety against all threats. He/she wants the state to be strong so it can defend its citizens. 15 He/she looks for adventures and likes to take risks. He/she wants to have an exciting life. 16 It is important to him/her to always behave properly. He/she wants to avoid doing anything people would say is wrong. 17 It is important to him/her to get respect from others. He/she wants people to do what he/she says. 18 It is important to him/her to be loyal to his/her friends. He/she wants to devote himself/herself to people close to him/her. 19 He/she strongly believes that people should care for nature. Looking after the environment is important to him/her. 20 Tradition is important to him/her. He/she tries to follow the customs handed down by his/her religion or his/her family. 21 He/she seeks every chance he/she can to have fun. It is important to him/her to do things that give him/her pleasure. 22 Religion plays an important role in his/her life. He/She tried to live up to his/her destiny. 23 He/She works hard, conscientiously and persistently. Punctuality and order are typical for him/her.	
35	How many times have you taken part in research on decision-making before?	Open question
36	Which is the highest level of education that your father achieved?	O Primary school

		O Secondary school O High school O Undergraduate degree O Master O Ph.D.
37	Which is the highest level of education that your mother achieved?	O Primary school O Secondary school O High school O Undergraduate degree O Master O Ph.D.
38	Which is your father's current job?	Open question
39	Which is your mother's current job?	Open question
40	Please write your household's yearly income, including all salaries, pensions, and other returns, net of taxes and other deductions.	Open question
41	Please write below your motivations for the decisions that you made during this research.	Open question
42	Please write below if you wish your opinions on this research.	Open question
43	In which city do you think the other lab was located?	Open question

S7 List of Abbreviations

APLA = Average PLA = Average Probability of Loss Avoidance.

AS = Anti-social Sanctioning - instances in which an *ego* punished an *alter* who contributed no less than the group median.

B-treatments = Blind Treatments: Participants were not made aware that students from the other laboratory were actually from another country.

C = Total contributions by a group.

c_i = Individual contribution

c_{-i} = Strategy profile of the other players except *i*. $c_{-i} = (c_1, ..., c_{i-1}, c_{i+1}, ..., c_n)$

CRSD = Collective Risk Social Dilemma.

CS = Cooperative Solution: It takes the perspective of the entire group and maximises the total sum of expected monetary payoffs.

GER_NAT_NS = Within-country treatment in Germany without sanctions.

GER_NAT_S = Within-country treatment in Germany with sanctions.

INT = International Level of Interaction.

INT_Blind_NS = International (between-countries) treatment without sanctions with "blind" interaction when participants did not know that they were interacting with people from another country.

INT_Blind_S = International (between-countries) treatment with sanctions with "blind" interaction when participants did not know that they were interacting with people from another country.

INT_Open_NS = International (between-countries) treatment without sanctions with "open" interaction when participants knew that they were interacting with people from another country.

INT_Open_S = International (between-countries) treatment with sanctions with "open" interaction when participants knew that they were interacting with people from another country.

KW = Kruskal-Wallis test.

L = Percentage of a loss to each player's private account if $C < T$.

n = Number of persons in the group (in our experiment n=6).

NAT = National Level of Interaction (within Germany (GER) or Russia (RUS)).

NE = Nash Equilibrium: It identifies the set of individual actions such that each action is the best response to others' individual actions, assuming that each agent maximises his/her own monetary payoff.

NS-treatments = Treatments without sanctions.

O-treatments = Open Treatments: German and Russian participants were informed that the other city was located either in Russia or in Germany, respectively.

P = Probability of Loss Avoidance.

PLA = Probability of Loss Avoidance.

PS = Pro-social Sanctioning - instances in which an *ego* punished an *alter* who contributed less than the group median.

RUS_NAT_NS = Within-country treatment in Russia without sanctions.

RUS_NAT_S = Within-country treatment in Russia with sanctions.

s = Share of the private account that is not lost in case the loss event occurs; $s = 1 - L$

$Sanction_{t-1}$ = Variable identifying whether a participant had been sanctioned in the previous period.

S-treatments = Treatments with sanctions included at the second stage of each period of the experiment.

ESM = Electronic Supplementary Materials.

T = Certain safety thresholds equal to 2100 tokens.

Tokens_Deducted(t-1) = Amount of tokens deducted from a participant's account in the previous period.

w = Initial endowment in each period, equal to 60 tokens in NS-treatments or 50 tokens at the first stage plus 10 tokens at the second stage in S-treatments.

WMW = Wilcoxon-Mann-Whitney tests.

ΔCooperation = Difference in Contribution to the collective fund between the current Period and the previous Period.

ESM References

1. Barrett, S. 2007 *Why cooperate? The incentive to supply global public goods*. Oxford, UK: Oxford University Press.

2. Nowak M, Sigmund, K. 2005 Evolution of indirect reciprocity. *Nature* **437** (7063), 1291-1298.

3. Handley, C, Mathew, S. 2020 Human large-scale cooperation as a product of competition between cultural groups. *Nature Comms.* **11**, 1-9.

4. Choi, J, Bowles, S. 2007 The coevolution of parochial altruism and war. *Science* **318***,* 636-640.

5. Bernhard, H, Fischbacher, U, Fehr, E. 2006 Parochial altruism in humans. *Nature* **442** (7105), 912-915.

6. Romano, A, Sutter, M, Liu, JH, Yamagishi, T, Balliet, D. 2021 National parochialism is ubiquitous across 42 nations around the world. *Nature Comms.* **12**, 1-8.

7. Dorrough, AR, Glöckner, A. 2016 Multinational investigation of cross-societal cooperation. *P. Natl. Acad. Sci. U.S.A.* **113**, 10836-10841.

8. Martin, LL. 1994 *Coercive cooperation: Explaining multilateral economic sanctions*. Princeton USA: Princeton University Press.

9. Keohane, R, Victor, D. 2016 Cooperation and discord in global climate policy. *Nat. Clim. Change* **6**, 570–575.

10. Przepiorka, W, Diekmann, A. 2013 Individual heterogeneity and costly punishment: a volunteer's dilemma. *Proc. Royal Soc. B* **280** (1759), 20130247.

11. Henrich, J, McElreath, R, Barr, A, Ensminger, J, Barrett, C, Bolyanatz, A,...

Lesorogol, C. 2006 Costly punishment across human societies. *Science* **312**, 1767-1770.

12. Boyd, R, Gintis, H, Bowles, S, Richerson, PJ. 2003 The evolution of altruistic punishment. *P. Natl. Acad. Sci. U.S.A.* **100**, 3531–3535.

13. Gürerk, Ö, Irlenbusch, B, Rockenbach, B. 2006 The Competitive Advantage of Sanctioning Institutions. *Science* **312**, 5770, 108-111. (doi:10.1126/science.1123633)

14. IPCC. 2021 *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [Masson-Delmotte, V, Zhai, P, Pirani, A, Connors, S L, Péan, C, Berger, S, Caud, N, Chen, Y, Goldfarb, L, Gomis, MI, Huang, M, Leitzell, K, Lonnoy, E, Matthews, JBR, Maycock, TK, Waterfield, T, Yelekçi, O, Yu, R, Zhou, B (eds.)]. Cambridge, UK: Cambridge University Press.

15. Nordhaus, W. 2015 Climate clubs: Overcoming free-riding in international climate policy. *Am. Econ. Rev.* **105**, 1339-70.

16. Berger, A, Brandi, C, Bruhn, D 2017 *Environmental provisions in trade agreements: promises at the trade and environment interface.* Bonn, Germany: German Development Institute, Briefing Paper, **16**.

17. Milinski, M, Sommerfeld, R, Krambeck, HJ, Reed, F, Marotzke, J. 2008 The collective-risk social dilemma and the prevention of simulated dangerous climate change. *P. Natl. Acad. Sci. U.S.A.* **105**, 2291-2294.

18. Alexander, M, Christia, F. 2011 Context modularity of human altruism. *Science* **334**, 1392-1394.

19. Gächter, S, Herrmann, B. 2009 Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment. *Phil. Trans. R. Soc. B* **364**, 791–806.

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

21. Kriegler, E, Hall, JW, Held, H, Dawson, R, Schellnhuber HJ. 2009 Imprecise probability assessment of tipping points in the climate system. *P. Natl. Acad. Sci. U.S.A.* **106**, 5041-5046.

22. Barrett, S, Dannenberg, A. 2012 Climate negotiations under scientific uncertainty. *P. Natl. Acad. Sci. U.S.A.* **109**, 17372-17376.

23. Finocchiaro Castro, M. 2008 Where are you from? Cultural differences in public good experiments. *J. Socio-Econ.* **37**, 2319-2329.

24. Böhm, R, Rockenbach, B. 2013 The inter-group comparison–intra-group cooperation hypothesis: comparisons between groups increase efficiency in public goods provision. *PLoS ONE* **8**, e56152.

25. Nikiforakis, N. 2008 Punishment and counter-punishment in public good games: can we really govern ourselves? *J. Publ. Econ.* **92**, 91–112.

26. Ostrom, E, Burger, J, Field, CB, Norgaard, RB, Policansky, D. 1999 Revisiting the commons: local lessons, global challenges. *Science* **284**, 278-282.

27. Inglehart, R, Welzel, C. 2005 *Modernization, Cultural Change and Democracy* New York, USA: Cambridge University Press.

28. Gächter, S, Herrmann, B, Thöni, C. 2010 Culture and cooperation. *Phil. Trans. R. Soc. B* **365**, 2651-2661.

29. Gächter, S, Herrmann, B. 2011 The limits of self-governance when cooperators get punished: Experimental evidence from urban and rural Russia. *Eur. Econ. Rev*. **55**, 193-210.

30. Balliet, D, Wu, J, De Dreu, CK. 2014. Ingroup favoritism in cooperation: a meta-analysis. *Psychol*. *Bull*. **140**,1556–1581.

31. Gino, F, Ayal, S, Ariely, D. 2009 Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. *Psychol. Sci.* **20**, 393-398.

32. Fiedler, S, Hellmann, DM, Dorrough, AR, Glöckner, A. 2018 Cross-national in-group favoritism in prosocial behavior: Evidence from Latin and North America. *Judgm. Decis. Mak.* **13**, 42–60.

33. Carpenter, J, Cardenas, JC. 2011 An intercultural examination of cooperation in the commons. *J. Conflict Resolut.* **55**, 632-651.

34. Goerg, SJ, Hennig-Schmidt, H, Walkowitz, G, Winter, E. 2016 In wrong anticipation. Miscalibrated beliefs between Germans, Israelis, and Palestinians. *PLoS ONE* **11**, e0156998.

35. Schmidt, KM, Ockenfels, A. 2021 Focusing climate negotiations on a uniform common commitment can promote cooperation. *P. Natl. Acad. Sci. U.S.A.* **118,** e2013070118. (doi:10.1073/pnas.2013070118)

36. Tavoni, A, Dannenberg, A, Kallis, G, Löschel, A. 2011 Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *P. Natl. Acad. Sci. U.S.A.* **108**, 11825-11829.

37. Hardin, G. 1968 The tragedy of the commons. *Science* **162**, 1243-1248.

38. Jacquet, J, Hagel, K, Hauert, C, Marotzke, J, Röhl, T, Milinski, M. 2013 Intra-and

intergenerational discounting in the climate game. *Nat. Clim. Change* **3**, 1025-1028.

39. Froehlich, L, Dorrough, AR, Glöckner, A, Stürmer, S. 2021 Similarity Predicts Cross-National Social Preferences. *Soc. Psychol. Personal. Sci.* **12**, 1486-1498.

40. DeCremer, D, Van Vugt, M 1999 Social identification effects in social dilemmas: A transformation of motives. *Europ. J. Soc. Psychol.* **29**, 871-893.

41. Ispano, A, Schwardmann, P. 2017 Cooperating over losses and competing over gains: A social dilemma experiment, *Game. Econ. Behav.* **105**, 329-348.

42. Dorrough AR, Froehlich L, Eriksson K. 2021 Cooperation in the cross-national context. *Curr Opin Psychol*. **44**, 281-285. (doi: 10.1016/j.copsyc.2021.10.006)

43. Fischbacher, U, Gächter, S, Fehr, E. 2001 Are people conditionally cooperative? Evidence from a public goods experiment. *Econ. Lett*. **71**, 397-404.

44. Hovi, J, Ward, H, Grundig, F. 2015 Hope or despair? Formal models of climate cooperation. *Environ. Res. Econ.* **62**, 665-88.

45. Gächter, S, Renner, E. 2018 Leaders as role models and 'belief managers' in social dilemmas. *J. Econ. Behav. Org*. **154**, 321- 334.

46. Muthukrishna, M, Bell, AV, Henrich, J, Curtin, CM, 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. *Psychol. Sci.* **31**, 678-701.

47. Al-Ubaydli, O, List, JA, Suskind, DL. 2017 What can we learn from experiments? Understanding the threats to the scalability of experimental results. *Am. Econ Rev.* **107**, 282-286.

48. Messner, D, Guarín, A, Haun, D. 2015 The behavioral dimensions of international cooperation. In *Global Cooperation and the Human Factor in International Relations* (eds. D. Messner, S. Weinlich), Abingdon, UK:Routledge) pp. 65-83.

49. Guala, F. 2012 Reciprocity: weak or strong? What punishment experiments do (and do not) demonstrate. *Behav. Brain Sci.* **35**, 1-15.

50. European Commission (2021): *Proposal for a Regulation of The European Parliament and of The Council establishing a carbon border adjustment mechanism*, Brussels, 14.7.2021. COM(2021) 564 final, 2021/0214 (COD), Document 52021PC0564. (https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52021PC0564; accessed 2022/01/18)

51. OECD. 2017 *Education at a Glance 2017: OECD Indicators,* Paris: OECD Publishing.

52. World Values Survey: http://www.worldvaluessurvey.org/.

53. Hofstede, G. 2011 Dimensionalizing cultures: The Hofstede model in context. *Online readings in psychology and culture* **2**, 2307-0919. (doi 10.9707/2307-0919.1014)

54. Hofstede insights: https://www.hofstede-insights.com/product/compare-countries/

55. Falk, A, Becker, A, Dohmen, T, Enke, B, Huffman, D, Sunde, U. 2018 Global evidence on economic preferences. *Q. J. Econ.* **133**, 1645–1692.

56. Weber, TO, Weisel, O, Gächter, S. 2018 Dispositional free riders do not free ride on punishment. *Nat. Comm.* **9,** 2390, 1-9.

57. Abadie, A, Athey, S, Imbens, GW, Wooldridge, J. 2017 *When should you adjust standard errors for clustering?* National Bureau of Economic Research, WP Series 24003. (doi 10.3386/w24003)

58. Kim, DG. 2020 *Clustering Standard Errors at the "Session" Level*, CESifo Working Paper No. 8386.

59. Moffatt, PG. 2016 *Experimetrics: Econometrics for experimental economics*, London, New York: Palgrave, Macmillan International Higher Education.

60. Sylwester, K, Herrmann, B, Bryson, JJ. 2013 Homo homini lupus? Explaining antisocial punishment. *J. Neurosci. Psychol. E*. **6**, 167-188.

61. Dreber, A, Rand, DG, Fudenberg, D, Nowak, MA. 2008 Winners don't punish. *Nature* **452**, 348-351

62. Gächter, S, Renner, E, Sefton, M. 2008 The long-run benefits of punishment. *Science* **322**, 1510-1510.

63. Buchan, NR, Brewer, MB, Grimalda, G, Wilson, RK, Fatas, E, Foddy, M. 2011 Global social identity and global cooperation. *Psychol. Sci.* **22**, 821-828.

64. Inglehart, R, Haerpfer, C, Moreno, A, Welzel, C, Kizilova, K, Diez-Medrano, J, Lagos, M, Norris, P, Ponarin, E, Puranen, B, et al. (eds.) 2018 *World Values Survey: Round Six - Country-Pooled Datafile*. Madrid, Vienna: JD Systems Institute & WVSA Secretariat. (doi.org/10.14281/18241.8)

65. Fischbacher, U. 2007 z-Tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ.* **10**, 171-178.

66 Roth, AE, Prasnikar, V, Okuno-Fujiwara, M, Zamir, S. 1991 Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental study. *Am. Econ Rev*. **81**, 1068-1095.

67. Costa, A, Foucart, A, Arnon, I, Aparici, M, Apesteguia, J. 2014 ''Piensa'' Twice: On the foreign language effect in decision making. *Cognition* **130**, 236–254.

68. Snowberg, E, Yariv, L. 2021 Testing the waters: Behavior across participant pools. *Am. Econ Rev*. **111**, 687-719.

69. Levitt, SD, List, JA. 2007 What do laboratory experiments measuring social preferences reveal about the real world? *J. Econ. Perspect.* **21**, 153-174.

70. Galizzi, MM, Navarro-Martinez, D. 2019 On the external validity of social preference games: a systematic lab-field study. *Manage. Sci.* **65**, 976-1002.

71. Falk, A, Heckman JJ. 2009 Lab experiments are a major source of knowledge in

social sciences. *Science* **326**, 535-538.

72. Cappelen, AW, Nygaard, K, Sørensen, EØ, Tungodden, B. 2015 Social preferences in the lab: A comparison of students and a representative population. *Scand. J. Econ*. **117**, 1306-1326.

73. Falk, A, Meier, S, Zehnder, C. 2013 Do lab experiments misrepresent social preferences? The case of self-selected student samples. *J. Eur. Econ. Assoc*. **11**, 839-852.

74. Bock, O, Baetge, I, Nicklisch, A. 2014 hroot: Hamburg Registration and Organization Online Tool, *Eur. Econ. Rev.* **71**, 117-120.

75. Eckel, CC, Wilson, RK. 2006 Internet cautions: Experimental games with internet partners. *Exp. Econ.* **9**, 53–66.

76. Schwartz, SH. 1992 Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Adv. Exp. Soc. Psychol.* **25**, 1–65.

S8 Acknowledgements

We would like to thank all assistants that have helped us throughout our research: Anaida Arakelyan; Margarita Buchholz; Anja Bodenschatz; Ekaterina Chegaeva; Gregory Chernov; Ekaterina Goncharova; Alexei Guzey; Anne Kielmann; Johannes Marzian; Henrik Mertens; Thore Petersen; Aleksandra Shchukina; Felix Steinle; Patricia Trapp; Andrei Zaloilo; Dmitry Zhmurovsky; Olga Zykina. Thore Petersen programmed the Z-tree software we used to collect data.

We also acknowledge funding from the Center for Global Cooperation Research at the University of Duisburg-Essen, Christian-Albrechts-University of Kiel, and the National Research University Higher School of Economics- within the framework of the Basic Research Program at the National Research University Higher School of Economics (HSE).