aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Perspectives on Psychological Science
8(4) 363–379

DOI: 10.1177/1745691613488533
pps.sagepub.com
SAGE

Trust, Punishment, and Cooperation Across 18 Societies: A Meta-Analysis

Daniel Balliet and Paul A. M. Van Lange
Department of Social and Organizational Psychology, VU University Amsterdam, the Netherlands

Abstract
Punishment promotes contributions to public goods, but recent evidence suggests that its effectiveness varies across societies. Prior theorizing suggests that cross-societal differences in trust play a key role in determining the effectiveness of punishment, as a form of social norm enforcement, to promote cooperation. One line of reasoning is that punishment promotes cooperation in low-trust societies, primarily because people in such societies expect their fellow members to contribute only if there are strong incentives to do so. Yet another line of reasoning is that high trust makes punishment work, presumably because in high-trust societies people may count on each other to make contributions to public goods and also enforce norm violations by punishing free riders. This poses a puzzle of punishment: Is punishment more effective in promoting cooperation in high- or low-trust societies? In the present article, we examine this puzzle of punishment in a quantitative review of 83 studies involving 7,361 participants across 18 societies that examine the impact of punishment on cooperation in a public goods dilemma. The findings provide a clear answer: Punishment more strongly promotes cooperation in societies with high trust rather than low trust.

Keywords
trust, culture, punishment, cooperation, social capital, meta-analysis

One pervasive challenge to groups, organizations, and societies is the free-rider problem: People may benefit from collectively provided public goods but not contribute their fair share. The term *free riding* may well have its roots in someone using public transportation without paying the fare—and thus enjoying a "free ride." Although tempting, the quality of the public good in question, public transportation in this case, might be undermined as there are more people free riding. Likewise, it is in the individual's interest to minimize paying taxes and keep as much as possible for oneself, yet the nation needs income from taxes to optimally serve its members—by maintaining the quality of public goods, such as education, infrastructure, and safety in society. Or in the context of smaller groups, just imagine a group project where the quality of the project is enhanced by the contributions of each group member. However, it may be tempting for each individual to allocate valuable time and energy to more enjoyable activities, while hoping that the other group members will work on the group project.

Over the past several decades, social and behavioral scientists have examined behavior in such public good dilemmas with the use of an experimental paradigm in which participants are faced with a decision about how much to contribute (e.g., money) to a public good. Since its beginning, one of the key questions has been "How do groups and societies minimize free riding and instead promote contributions to public goods?". Several solutions have been examined, and to some degree supported, in past research. For example, people are less likely to free ride and more likely to cooperate if people communicate, if they perceive their cooperative efforts to be more efficacious, and when people are less anonymous (for reviews, see Balliet, 2010; Chaudhuri, 2011; Gächter & Herrmann, 2009; Van Lange, Balliet, Parks, & Van Vugt, in press). A potentially even more powerful solution has been suggested by recent research: Free riding is dramatically reduced if people in a public good task are endowed with the option to punish each other.

Corresponding Author:
Daniel Balliet, Department of Social and Organizational Psychology, VU University, Van der Boechorststraat 1, 1081 BT, Amsterdam, the Netherlands
E-mail: d.p.balliet@vu.nl

Indeed, in two seminal articles, Fehr and Gächter (2000, 2002) developed a public good paradigm that demonstrated that the use of punishment could effectively increase contributions to public goods. In this paradigm, participants interact in a small group and have an opportunity to contribute to a public good or to free ride on the contributions of their group members. After making their decision, each group member learns about the decision of the other group members, and at that point they decide to pay a cost to reduce the earnings of any specific group member(s)—thus punishing their group member(s). Fehr and Gächter found that over time people tended to make greater contributions to the public good with the presence of punishment opportunities than they did when punishment opportunities were absent. Subsequent research replicated and supported the conclusion that individuals are willing to pay a cost to punish free riders and that this can effectively and efficiently increase contributions to public goods (Gächter, Renner, & Sefton, 2008; Gürerk, Irlenbusch, & Rockenbach, 2006; Ostrom, Walker, & Gardner, 1992; Rand, Dreber, Ellingsen, Fudenberg, & Nowak, 2009; Rockenbach & Milinski, 2006; Yamagishi, 1986; for a recent meta-analysis, see Balliet, Mulder, & Van Lange, 2011).[1]

Recently, cross-societal studies on punishment have found much variation across societies in the degree to which punishment promotes cooperation (Balliet et al., 2011; Henrich, Ensminger, et al., 2010; Henrich et al., 2006; Herrmann, Thöni, & Gächter, 2008; Marlowe & Berbesque, 2008). In many societies, punishment is effectively directed toward free-riders and promotes cooperation; yet in other societies, the effectiveness of punishment to promote contributions to public goods is weak or even absent (Herrmann et al., 2008). For example, in some societies, such as Denmark and China, punishment works well, whereas in other societies, such as Turkey and South Africa, punishment hardly works at all. One important reason punishment works is because punishment is largely delivered to free riders. However, punishment does not work in some societies because people punish cooperators or group members who contribute more than themselves—so-called "antisocial punishment" (Herrmann et al., 2008). Certainly, the punishment of cooperators can undermine cooperation. Thus, in some societies, punishment can be countereffective.

Can some aspect of cultural differences explain why the effectiveness of punishment in promoting cooperation is so strongly dependent on the societies in which the studies are conducted? Clearly, societies differ in a number of ways, but one key cultural difference among various societies is the beliefs that people hold about other people's benevolence—that is, in trust (Bond et al., 2004). Trust is often defined in terms of the willingness to accept vulnerability on the basis of the positive expectations of the intentions or behaviors of others (Rousseau, Sitkin, Burt, & Camerer, 1998; see also Balliet & Van Lange, 2012). As shown repeatedly, individuals differ in how much they trust others in general (Rotter, 1967; Yamagishi, 1988, 2011), groups and organizations differ in the extent to which group members trust each other (Dirks, 1999), and even societies can differ in the degree to which societal members trust others (Inglehart, Basanez, & Moreno, 1998; Knack & Keefer, 1997). For example, in some societies a lost wallet is believed to be returned immediately, whereas in other societies people are not that trusting (Knack, 2001). Such differences in trust are important to public goods, because people with high trust are more likely to contribute (Balliet & Van Lange, 2012; Fischbacher, Gächter, & Fehr, 2001; Gächter, Herrmann, & Thöni, 2004; Yamagishi, 1988) and varying levels of trust might reflect norms about expected contributions to public goods (Coleman, 1988; Doney, Cannon, & Mullen, 1998). In the present research, we examine the possibility that differences in trust among societies play a key role for understanding the effectiveness of punishment to promote cooperation.

How might punishment work in societies that differ in trust? One perspective assumes that punishment should be especially effective in low-trust societies. The reasoning is that in high-trust societies, most people already contribute to the public good: The belief that most others can be trusted renders punishment an unnecessary cost to encourage contributions. Also, in high-trust societies, people may believe that most people have internalized norms against free riding, so that norm enforcement through punishment is not needed. In contrast, in low-trust societies, punishment is necessary because people do not expect their fellow members to contribute in the absence of incentives to do so, but in low-trust societies the presence of punishment increases expectations that others will contribute to public goods (Yamagishi, 1986, 1988).

Another, competitive perspective assumes that punishment is especially effective in high-trust societies. The reasoning is that in high-trust societies, societal members adhere to norms that encourage both cooperation and punishment of free riders. Specifically, punishment of free riders poses a "second-order" dilemma, such that it is in each group member's material interest to not punish norm violators and to free ride on the benefits provided by others who do punish. Thus, in high-trust societies people may count on each other to enforce norm violations and punish free riders (Coleman, 1988). Moreover, in high-trust societies, when people deviate from these norms and free ride on others' contributions, punishment may induce a sense of guilt and encourage subsequent

cooperation. However, in low-trust societies, these norms might be less strongly shared and enforced, such that punishment may induce anger and fail to encourage cooperation, resulting in the antisocial punishment observed in certain societies (Gintis, 2008).

This poses the following puzzle of punishment: Does punishment promote greater contributions to public goods in low or high-trusting societies? We sought to illuminate this puzzle of punishment by conducting a quantitative analysis of experimental studies implemented across 18 societies on the effect of punishment on contributions in a public goods dilemma.

Additional Issues: Market Competitiveness, Religion, and Social Capital

Besides trust, we were able to test some other possible reasons why punishment would be more or less effective in promoting cooperation across societies. Prior research on the use of punishment to promote cooperation across small-scale societies has found that punishment is more frequently used to sustain cooperation in societies with a strong market economy or with a greater number of societal members adopting a world religion (Henrich, Ensminger, et al., 2010). Although the present analysis includes only large-scale societies that all possess market economies and exposure to world religions, these societies do vary in the functioning of their markets and the extent to which societal members participate in religion. Therefore, we also examined whether variation across these societies in market competitiveness and participation in religion predicted the effectiveness of punishment to promote cooperation.

Theory of social capital also applies to the present analysis. *Social capital* is essentially the idea that the benevolence of others in one's social network is a valuable resource that provides benefits (and sometimes costs) to individuals in terms of both economic and social exchange (Adler & Kwon, 2002). Two facets of social capital include trust and cooperative norms accompanied by sanctions (Coleman, 1988; Ostrom & Ahn, 2008; Portes, 1998). Therefore, this perspective predicts a positive correlation across societies among trust, cooperative norms, and the effectiveness of punishment in promoting cooperation. Thus, this perspective similarly predicts that cross-societal variation in trust will positively relate to the effectiveness of punishment to promote cooperation. Additionally, this perspective predicts that societies that possess stronger norms for cooperation will also display greater effectiveness of punishment. We will test this perspective by examining whether variation in norms of civic cooperation relates to the effectiveness of punishment.

Overview of the Meta-Analysis

In this article, we report a meta-analysis involving 83 studies of the impact of punishment on cooperation across 18 different countries, involving a total number of 7,361 participants. To maximize the consistency across studies, we included only those studies that used the experimental paradigm developed by Fehr and Gächter (2002). All studies included in this meta-analysis used this protocol. By applying a meta-analytic approach, we aim to explain variation in the effect size by considering cross-societal differences in trust while controlling for several methodological study characteristics that vary across studies as well as between-country differences in wealth, wealth inequality, democracy, market competitiveness, religious participation, and norms of civic cooperation.

Methods

Search for studies and study criteria

We searched several databases for published studies, including PsycINFO, EconLit, Google Scholar, and the Web of Science. We searched the entire text of English-language journal articles by using the term *punishment* along with *cooperation* terms (e.g., *public goods dilemmas* and *voluntary contribution mechanism*). We searched the references of all relevant articles for studies. We also contacted over 150 researchers who attended the 2007 and 2009 International Conference for Social Dilemmas for data. In addition, we posted a call for data to the Economic Science Association methods group (http://groups.google.com/group/esa-discuss). Last, we requested data from authors who published articles on punishment in public good dilemmas in the last 5 years.

All studies had to have been conducted on adult participants (age 18 and above). Studies had to have examined the effect of punishment on cooperation in a public goods dilemma. We included only those studies that compared a punishment condition with a control condition. As noted earlier, we included only those studies that manipulated peer punishment by using the methods developed by Fehr and Gächter (2000, 2002). In this paradigm, several people come to the lab and are randomly assigned to play a public goods dilemma game with three other persons.

The typical public goods experiment proceeds in two stages. In the first stage, participants make their decision about how much to contribute to the public good. Each participant is endowed 10 monetary units (MUs) and then decides between how much of that money to donate to a group fund (the public good) or an individual fund. Any amount contributed to the group fund is multiplied by 1.5, and then distributed equally across all group

members. The amount allocated to the individual fund is not multiplied. Each group member simultaneously and confidentially makes his or her decision. Next, the second stage begins. Participants are made aware of their group members' decisions and are given an opportunity to pay to punish their group members. Specifically, participants are given the opportunity to pay 1 MU to reduce their partner's earnings by 3 MUs. Last, participants are told whether they were punished, and how much they earned on that trial. The game is often played for several trials—either with the same group members (partner design) or with random assignment to a new group on each trial (stranger design). The effect of punishment on contributions is derived by comparing a punishment condition with a control condition, where participants make only their contribution decisions and are not provided the opportunity to punish others. Using these criteria to find relevant studies uncovered a total of 33 papers that contained 83 studies (effect sizes) and 7,361 participants.

Coding procedure

Although the studies included in the meta-analysis involved a highly standardized experimental protocol, the studies varied by the number of group members, number of iterations, and whether the participants continued to remain in the same group throughout the experiment or were reassigned to a new group for each trial. We coded these study characteristics, as described below, and controlled for these between-study differences when conducting our analyses. We used the same coding criteria as employed in prior research (Balliet et al., 2011). Specifically, we coded group size, the number of iterations, and whether the study employed a partner-matching or stranger-matching protocol. We also coded cross-societal variation in wealth, wealth inequality, market competitiveness, religious participation, norms of civic cooperation, the extent of democracy, and generalized trust in others, which we describe below.

Partner protocol. All studies in this meta-analysis used the paradigm developed by Fehr and Gächter (2000, 2002). Fehr and Gächter had participants play an iterated 4-person public goods dilemma game allowing participants an opportunity to punish their partners in the dilemma after each trial. Originally, Fehr and Gächter composed two conditions: the stranger design and partner design. In the stranger design, participants play out the same dilemma for several trials but are randomly assigned to a new group after each trial. In the partner design, however, participants remain anonymous but are not reassigned to a new group after each trial and remain in the same group for the entire experiment. We coded whether the study used the partner design (k = 45) or stranger design (k = 28).

Number of iterations. Participants were allowed to interact only once, or the dilemma could occur repeatedly for several iterations. The sample of effect sizes includes both one-shot (k = 9) and iterated (k = 72) dilemmas. We coded the number of iterations as a continuous variable (M = 10.24, Mdn = 10).

Group size. We coded group size as a continuous variable (range = 3 to 7). The mode of group size was a four-person group (k = 46; M = 4, Mdn = 4).

Country of participants. The studies were conducted in 18 different countries. Most studies were conducted in the United States (k = 15), United Kingdom (k = 15), Switzerland (k = 10), Russia (k = 7), Netherlands (k = 6), and Israel (k = 4). Other countries represented include Australia, Belarus, China, Denmark, Germany, Greece, Italy, Saudi Arabia, South Africa, South Korea, Turkey, and Ukraine. The country for each effect size is labeled in Table 1.

National wealth. We used gross domestic product (GDP) per capita as an index of a country's wealth. We coded GDP per capita for each society by using the index provided by the International Monetary Fund (http://www.imf.org/external/pubs/ft/weo/2011/02/weodata/index.aspx). We used the estimates of GDP per capita for the year 2011. Because this variable is not normally distributed, we transformed GDP by using the logarithm. All analyses were conducted with the transformed logarithmic GDP variable. Higher scores indicate more wealthy countries, and the logarithmic scores ranged from 8.19 for Ukraine to 11.30 for Switzerland (M = 10.48, SD = 0.70).

Wealth inequality. The Gini index is used as a measure of cross-societal differences in wealth inequality. We used data reported in the Human Development Report (United Nations Development Programme, 2010). Of the societies represented in our sample the Gini values range from 24.7 for Denmark to 67.4 for South Africa (M = 36.17, SD = 9.52). Lower Gini scores indicate greater amounts of equality in wealth.

Market competition. We coded each country's score on the global competitive index reported by the World Economic Forum (Schwab, 2012). This is a measure of a country's competitiveness and productivity. Competiveness is defined as a "set of institutions, policies, and factors that determine the level of productivity in a country" (Schwab, 2012, p. 4). The index considers 12 different

Table 1. Experiments Included in the Meta-Analysis That Report the Effect of Punishment on Cooperation in a Public Goods Dilemma

Study	*N*	Country	*d*	95% CI	SD or PM	IT(#)	GS
Bochet et al. (2006)	116	US	0.96	[0.57, 1.70]	PM	IT(10)	4
Bornstein & Weisel (2010)							
Sample a	72	IL	1.42	[0.76, 2.07]	PM	IT(18)	4
Sample b	72	IL	0.52	[0.03, 1.01]	PM	IT(18)	4
Bornstein & Weisel (2010)							
Sample a	72	IL	1.65	[0.94, 2.36]	PM	IT(18)	4
Sample b	72	IL	0.71	[0.20, 1.23]	PM	IT(18)	4
Carpenter (2007)							
Sample a	46	US	0.55	[0.14, 0.97]	SD	IT(10)	7
Sample b	46	US	0.44	[0.02, 0.85]	SD	IT(10)	7
Sample c	46	US	0.18	[–0.23, 0.59]	SD	IT(10)	7
Carpenter & Matthews (2009)	100	US	0.59	[0.19, 0.99]	PM	IT(10)	4
Carpenter et al. (2004)	72	US	0.94	[0.45, 1.44]	PM	IT(10)	4
Casari & Luini (2009)	60	IT	1.60	[0.30, 1.27]	PM	IT(10)	5
Egas & Riedl (2008)							
Sample a	324	NL	0.08	[–0.14, 0.31]	SD	IT(6)	3
Sample b	324	NL	0.50	[0.27, 0.72]	SD	IT(6)	3
Sample c	306	NL	0.21	[–0.01, 0.44]	SD	IT(6)	3
Sample d	324	NL	–0.09	[–0.31, 0.13]	SD	IT(6)	3
Ertan et al. (2009)	80	US	0.37	[0.14, 0.59]	PM	IT(3)	4
Fehr & Gächter (2000)							
Sample a	24	CH	1.35	[0.46, 2.23]	SD	IT(10)	4
Sample b	80	CH	1.65	[1.14, 2.15]	PM	IT(10)	4
Fehr & Gächter (2002)	236	CH	1.35	[1.17, 1.53]	SD	IT(10)	4
Fuster & Meier (2009)							
Sample a	15	US	1.10	[0.50, 1.74]	PM	IT(6)	4
Sample b	19	US	0.94	[0.40, 1.48]	PM	IT(6)	4
Gächter & Herrmann (2009)							
Sample a	141	CH	0.31	[0.14, 0.48]	SD	OS	3
Sample b	102	CH	0.30	[0.11, 0.50]	SD	OS	3
Sample c	180	RU	–0.05	[–0.20, 0.09]	SD	OS	3
Sample d	180	RU	–0.08	[–0.23, 0.06]	SD	OS	3
Gächter & Herrmann (2010)							
Sample a	205	RU	–0.07	[–0.21, 0.07]	SD	OS	3
Sample b	105	RU	0.07	[–0.13, 0.26]	SD	OS	3
Sample c	143	RU	0.18	[0.01, 0.36]	SD	OS	3
Sample d	153	RU	–0.17	[–0.33, –0.01]	SD	OS	3
Gächter et al. (2008)							
Sample a	105	UK	0.71	[0.31, 1.11]	PM	IT(10)	3
Sample b	102	UK	1.95	[1.48, 2.42]	PM	IT(50)	3
Gächter & Thoni (2005)							
Sample a	51	CH	1.85	[1.19, 2.50]	SD	IT(10)	3
Sample b	126	CH	0.20	[–0.15, 0.56]	SD	IT(10)	3
Sample c	54	CH	0.62	[0.07, 1.16]	SD	IT(10)	3
Herrmann et al. (2008)							
Sample a	56	US	1.36	[1.00, 1.72]	PM	IT(10)	4
Sample b	56	UK	1.38	[1.02, 1.75]	PM	IT(10)	4
Sample c	68	DK	1.12	[0.81, 1.42]	PM	IT(10)	4
Sample d	60	DE	1.42	[0.82, 1.47]	PM	IT(10)	4
Sample e	48	CH	1.87	[1.40, 2.34]	PM	IT(10)	4
Sample f	96	CH	1.21	[0.95, 1.48]	PM	IT(10)	4
Sample g	68	BY	0.55	[0.29, 0.80]	PM	IT(10)	4

(continued)

Table 1. (continued)

Study	*N*	Country	*d*	95% CI	SD or PM	IT(#)	GS
Sample h	44	UA	0.08	[–0.22, 0.38]	PM	IT(10)	4
Sample i	152	RU	0.39	[0.23, 0.56]	PM	IT(10)	4
Sample j	44	GR	–0.16	[–0.46, 0.14]	PM	IT(10)	4
Sample k	64	TR	0.33	[0.08, 0.58]	PM	IT(10)	4
Sample l	48	SA	–0.15	[–0.43, 0.14]	PM	IT(10)	4
Sample m	84	KR	1.64	[1.31, 1.97]	PM	IT(10)	4
Sample n	96	CN	1.23	[0.96, 1.49]	PM	IT(10)	4
Sample o	40	AU	2.14	[1.58, 2.71]	PM	IT(10)	4
Kocher et al. (2008)	120	ZA	0.22	[0.03, 0.40]	SD	OS	3
Kroll et al. (2007)	35	US	0.88	[0.49, 1.27]	PM	IT(10)	5
Myers (2009)	72	US	1.12	[0.63, 1.62]	SD	IT(20)	4
Nikiforakis (2008)							
Sample a	144	UK	1.03	[0.67, 1.40]	SD	IT(10)	4
Sample b	144	UK	1.63	[1.23, 2.04]	PM	IT(10)	4
Sample c	144	UK	0.21	[–0.14, 0.56]	SD	IT(10)	4
Sample d	144	UK	0.51	[0.16, 0.86]	PM	IT(10)	4
Nikiforakis & Normann (2008)							
Sample a	48	UK	0.50	[–0.07, 1.08]	PM	IT(10)	4
Sample b	48	UK	1.09	[0.49, 1.7]	PM	IT(10)	4
Sample c	48	UK	2.07	[1.37, 2.77]	PM	IT(10)	4
Sample d	48	UK	2.24	[1.52, 2.97]	PM	IT(10)	4
Nikiforakis et al. (2010)							
Sample a	44	UK	1.09	[0.45, 1.72]	PM	IT(10)	4
Sample b	48	UK	1.20	[0.58, 1.81]	PM	IT(10)	4
Sample c	44	UK	1.84	[1.14, 2.55]	PM	IT(10)	4
Sample d	48	UK	1.85	[1.18, 2.53]	PM	IT(10)	4
O'Gorman et al. (2008)							
Sample a	44	UK	0.35	[0.05, 0.66]	SD	IT(6)	4
Sample b	44	UK	–0.06	[–0.36, 0.24]	SD	IT(6)	4
O'Gorman & Van Vugt (2010)							
Sample a	48	UK	0.80	[0.45, 1.13]	SD	IT(6)	4
Sample b	44	UK	0.38	[0.08, 0.69]	SD	IT(6)	4
Page et al. (2005)	128	US	1.29	[0.91, 1.67]	PM	IT(20)	4
Patel et al. (2010)							
Sample a	66	UK	0.20	[–0.39, 0.80]	PM	IT(10)	4
Sample b	36	UK	0.38	[–0.45, 1.21]	PM	IT(10)	4
Sample c	72	UK	0.21	[–0.36, 0.78]	PM	IT(10)	4
Sample d	36	UK	–0.79	[–1.71, 0.13]	PM	IT(10)	4
Rand et al. (2009)	26	US	1.61	[0.71, 2.52]	PM	IT(50)	4
Reuben & Riedl (2009)							
Sample a	39	NL	4.51	[3.33, 5.70]	PM	IT(10)	3
Sample b	39	NL	0.90	[0.24, 1.56]	PM	IT(10)	3
Sample c	42	NL	2.38	[1.59, 3.17]	PM	IT(10)	3
Sample d	39	NL	1.77	[1.00, 2.54]	PM	IT(10)	3
Sefton et al. (2007)	18	US	0.32	[–0.15, 0.79]	PM	IT(10)	4
Tan (2008)	48	NL	1.33	[0.71, 1.96]	PM	IT(15)	4

Note: *N* = number of participants included in the effect size estimate; *d* = standardized mean difference; SD = stranger-matching design; PM = partner-matching design; IT(#) = Iterations (number of iterations); GS = group size; sample = a coded effect size; US = United States; IL = Israel; IT = Italy; NL = Netherlands; CH = Switzerland; RU = Russia; UK = United Kingdom; DK = Denmark; DE = Germany; BY = Belarus; UA = Ukraine; GR = Greece; TR = Turkey; SA = Saudi Arabia; KR = South Korea; CN = China; AU = Australia; ZA = South Africa.

components that are groupings of variables that impact a country's productivity. Of the societies included in our sample, scores ranged from 3.92 for Greece to 5.74 for Switzerland ($M = 5.20$, $SD = 0.50$). Higher scores indicate more competitive countries.

Religious participation. We used a measure of religious participation from the latest wave of the World Values Survey (WVS; http://www.worldvaluessurvey.org/) and the European Values Survey (EVS; http://www.europeanvaluesstudy.eu/).[2] We used the measure of church attendance over the last year with the question "apart from weddings, funerals, and christenings, about how often do you attend religious services these days?". Responses range from 1 (*more than once a week*) to 8 (*practically never*). Of the countries in our sample, church attendance ranged from 3.48 for South Africa to 6.20 for the Netherlands ($M = 5.61$, $SD = 0.53$). Higher scores indicate less religious participation.

Norms of civic cooperation. We used the latest wave of the WVS and EVS to measure norms of civic cooperation within each society. We followed previous research (e.g., Herrmann et al., 2008) by averaging responses to three items on the survey. Each item asks the respondent to what extent a specific behavior is justifiable: (a) claiming government benefits that you are not entitled to, (b) avoiding a fare on public transportation, and (c) cheating on taxes if you have a chance. Responses range from 1 (*never justifiable*) to 10 (*always justifiable*). We reverse-coded the scale so that high scores indicate greater norms of civic cooperation. In our sample, the index of norms of civic cooperation ranged from 7.00 for Greece to 9.39 for Turkey ($M = 8.70$, $SD = 0.50$). Higher scores equal stronger norms of civic cooperation.

Democracy. The extent of a society's democracy is taken from the democracy index. This index is based on several different measures of the electoral process, civil liberties, and political participation. We coded the democracy index from the Economist Intelligence Unit (2010). Of the countries in our sample, the democracy index ranged from 1.84 for Saudi Arabia to 9.52 for Denmark ($M = 6.91$, $SD = 2.29$). Higher scores indicate greater democracy.

Generalized trust in others. For each country we used the WVS and EVS to measure generalized trust in others. The WVS and EVS measure generalized trust with responses to the item "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?" (1 = *most people can be trusted* and 2 = *you can never be too careful when dealing with people*). To create a trust score for each society we examined observations from a country's latest wave of the WVS and EVS. The trust score indicates the percentage of people who claimed that most people can be trusted. Because the WVS is administered to people age 18 and above, and because the experimental studies included in the meta-analysis are primarily young adults (ages 18–25), we examined the correlation between age and trust within each society and coded the predicted score of trust for people age 21.5. Using this approach we have an estimated amount of trust in each society that better reflects the age group contained in the experimental studies used in the meta-analysis. To keep these predicted trust scores comparable to the trust index used in prior research we multiplied each predicted value by 2. Scores ranged from 8.60 for Turkey to 165.81 for Denmark ($M = 78.17$, $SD = 30.43$). Higher scores indicate greater amounts of trust.

Overview of analysis

In our analyses, we used the *d* statistic as the measure of effect size. This value is the difference between the mean levels of cooperation in the punishment and control conditions, divided by the pooled standard deviation. A positive *d* value shows greater cooperation in the punishment condition, relative to a control condition. We used the means and standard deviations to calculate the *d* values, but when these statistics were not reported, we calculated the *d* values by using the sample size along with the *F* score or *t* value. The *d* value is the dependent variable in our primary analyses.

Some studies allowed us to code multiple effect sizes. For example, a study could include a control condition and two different punishment conditions. This study allowed us to code two effect sizes. However, the two effect sizes are nonindependent because they share the same control condition. Therefore, we applied Cooper's (1998) shifting-units-of-analysis approach to handle nonindependent effect sizes when doing our analyses. Using this approach, we averaged over all the effects from a single study. This created one effect size for each study with multiple nonindependent effect sizes. These combined effect sizes were used in each of the analyses.

For the analyses we report below, we first estimate and report the overall average effect size for the effect of punishment on cooperation. We then report the average effect size in each of the countries represented in our sample. We calculated these estimated effect sizes by using a random effects model. A fixed-effect model is inappropriate because we assumed that we did not have the entire population of studies and that there is systematic between-study variation. Specifically, we assumed that the effect of punishment on cooperation would vary systematically across countries and that this variation may

be explained, at least in part, by cross-societal differences in the generalized trust in others. We assessed the estimated variation in the effect size distributions by using several indicators of heterogeneity of variance (T, T^2, and I^2).

We also examined the possibility that the effect size distribution contains a publication bias. In so doing, we formally examined the distribution of studies in a funnel plot (plotted according to their effect size and standard error) by using Egger's regression intercept and Duval and Tweedie's (2000) trim and fill approach.

Last, we report random effects multiple regression models with method of moments estimations by using the between-study and between-country moderators to predict the effect size. A random effects model is a relatively conservative statistical test compared with a fixed effect model (Lipsey & Wilson, 2001). To examine the hypothesis that cross-societal variation in trust relates to the effect size, we first tested the relation between trust and the effect size and then added several control variables to observe whether they changed this relation. In all of our models, because we had 69 observations, we included no more than seven predictor variables. To conduct our analyses we used the SPSS statistical MACROs provided by Lipsey and Wilson (2001).[3]

Results

Overall analysis of the effect of punishment on cooperation

We begin our analysis by reporting the main effect of punishment on cooperation across all societies. Table 1 reports the average effect of punishment on cooperation for each study included in the meta-analysis and lists the between-study codings of the moderators for each study. Prior to all analyses, we created an average effect size for studies with multiple nonindependent effect sizes. This reduced the sample of effect sizes (from $k = 83$ to $k = 69$).

Averaging across all societies, punishment had a moderate, positive effect on cooperation in the public goods dilemmas ($d = 0.77$, 95% CI [0.63, 0.91], *90%* prediction interval [0.16, 1.71]). Next, we assessed the estimated variation in the effect size distribution by using several indicators of heterogeneity of variance (T, T^2, and I^2). The indicators of heterogeneity of the effect size distribution suggest that there is variation in the true effect size distribution ($T^2 = .31$, $T = .55$) and that a substantial amount of this variation can be explained by between-study differences ($I^2 = 93.10\%$).

This effect size estimate could have resulted in a publication bias, as the majority of studies included in the analysis were published studies. In order to examine the possibility of a publication bias, we first considered the funnel plot where all studies were plotted according to their effect size and standard error (see online supporting materials at http://pps.sagepub.com/supplemental). We used the trim and fill approach to analyze symmetry in the funnel plot (Duval & Tweedie, 2000). This method examines the symmetry of the effect size in the funnel plot and then attempts to estimate where missing studies would fall in the plot that would produce symmetry in the funnel plot. The method estimates the value of these missing effects sizes through an iterative process of removing the most extreme effect sizes with small samples from either side of the plot. If there is a publication bias, then we can expect that this approach will estimate that the sample is missing studies below the overall effect size. The analysis then inputs these missing effect sizes to the sample of studies and calculates a smaller overall effect size without publication bias. In taking this approach, there were no estimated missing studies above the overall effect size. Asymmetry did exist, however, and was characterized by missing studies below the overall effect size. The trim and fill approach added 12 studies below the overall effect size and recalculated an average effect size that was slightly smaller than the original estimated effect size ($d = 0.57$, 95% CI [0.43,0.71]). Moreover, we found a statistically significant Egger's regression intercept (intercept = 4.68), $t(68) = 5.99$, $p < .001$, which indicates a possible publication bias. Taken together, these results suggest that the sample contained a possible publication bias that is usually characterized by missing null findings. Adjusting the estimated effect size while taking these studies into consideration resulted in a slightly smaller estimated effect of punishment on cooperation, compared with the overall analysis. An analysis of Orwin's fail-safe N finds that 546 studies with an effect size of $d = 0.00$ would be necessary to reduce the average effect size below $d = 0.05$.

Punishment and cooperation across 18 societies

Next, we report the effect sizes separately for each society. Table 2 reports the estimated average effect size for each society represented in the analysis along with the cross-societal codings of trust, national wealth, democracy, economic inequality, market competition, religious attendance, and norms of civic cooperation for each country. As displayed in Table 2, there is much cross-societal variation in the impact of punishment on cooperation. Although punishment effectively increases cooperation in some countries, such as Israel ($d = 1.27$), the Netherlands ($d = 1.76$), and Switzerland ($d = 1.04$), punishment is much less effective in other countries, such as Russia ($d = 0.03$), Greece ($d = -0.16$), and Saudi Arabia ($d = -0.15$). Next, we examined whether

Table 2. The Effect of Punishment on Cooperation Across 18 Societies and the Country Codings

Country	*k*	*d*	95 % CI	Trust	GDP	Gini	Dem.	Mkt.	Rel.	Norms
Australia	1	2.14	[1.58, 2.71]	93.80	11.09	35.19	9.22	5.11	6.07	8.95
Belarus	1	0.55	[0.29, 0.80]	48.15	8.68	27.22	3.34	—	5.68	7.49
China	1	1.23	[0.96, 1.49]	97.22	8.60	41.53	3.14	4.90	5.69	8.63
Denmark	1	1.12	[0.81, 1.41]	165.81	11.00	24.70	9.52	5.40	5.81	9.31
Germany	1	1.14	[0.82, 1.47]	81.07	10.69	28.31	8.38	5.41	6.00	8.84
Greece	1	–0.16	[–0.46, 0.14]	51.59	10.21	34.27	7.96	3.92	3.76	7.00
Israel	4	1.03	[0.51, 1.54]	49.51	10.37	39.20	7.48	5.07	—	—
Italy	1	1.60	[0.30, 2.90]	69.79	10.50	36.03	7.83	4.43	3.77	9.03
Netherlands	6	1.76	[0.72, 2.86]	106.16	10.83	30.90	8.99	5.41	6.20	9.09
Russia	7	0.03	[–0.11, 0.17]	56.58	9.47	42.27	4.26	4.21	5.99	7.80
Saudi Arabia	1	–0.15	[–0.43, 0.14]	106.86	9.93	32.00	1.84	5.17	4.33	8.32
South Africa	1	0.22	[0.03, 0.40]	33.86	9.00	67.40	7.79	4.34	3.48	8.41
South Korea	1	1.64	[1.31, 1.97]	74.50	10.03	31.51	8.11	5.02	4.56	8.63
Switzerland	10	1.04	[0.64, 1.44]	123.02	11.30	33.68	8.09	5.74	5.44	9.17
Turkey	1	0.33	[0.08, 0.58]	8.60	9.26	43.23	5.73	4.28	4.65	9.39
Ukraine	1	0.08	[–0.22, 0.38]	60.69	8.19	27.51	6.30	4.00	5.08	7.55
United Kingdom	15	0.70	[0.41, 0.99]	59.00	10.56	35.97	8.16	5.39	5.93	8.73
United States	15	0.81	[0.59, 1.02]	59.21	10.79	40.08	8.18	5.43	5.46	8.81

Note: *k* = number of effect sizes; *d* = standardized mean difference; *LL/UL* = Lower Limit/Upper Limit; GDP = the logarithm of the country's Gross Domestic Product per capita; GINI = Wealth inequality; Dem. = Democracy Index; Mkt = Market Competitiveness; Rel. = Religious attendance; Norms = Norms of civic cooperation.

cross-societal variation in trust may explain some of this cross-societal variation in the effectiveness of punishment to promote cooperation.

Moderator analyses

We were primarily interested in examining the relation between cross-societal variation in trust with the effect of punishment on cooperation in a public goods dilemma. When testing for this relation, we first examined the variance in the effect size distribution explained by cross-societal variation in trust (Model 1). We then observed how this relation was affected by adding the three study characteristics that vary across studies: partner protocol, number of iterations, and group size (Model 2) and then cross-societal differences in GDP per capita (Model 3), democracy (Model 4), and wealth inequality (Model 5). Next, because we restricted our model to no more than seven predictor variables at once (given a sample of 69 effect sizes), we dropped a nonsignificant cross-societal index from the model (wealth inequality) and then added either market competition (Model 6), religious attendance (Model 7), or norms of civic cooperation (Model 8).

We report the correlations between these variables and trust in Table 3. As displayed in Table 3, national wealth (GDP), wealth inequality (Gini), and market competitiveness all had significant correlations with trust. High-trust societies are wealthier, have less wealth inequality, and involve more market competition than do low-trust societies. Table 4 reports the results of the eight random effects multiple regression models using method of moments estimations.

As reported in Table 4, Model 1 explained a significant amount of variation in the effect size distribution (R^2 = .11, $p < .001$). Cross-societal variation in trust had a positive relation with the effect size (β = .33, p = .001). This indicates that punishment is more effective at increasing cooperation in high-trust societies than it is in low-trust societies.[4]

In Model 2, we added the study characteristics (i.e., partner protocol, number of iterations, and group size); this model explained a significant amount of variation in the effect size (R^2 = .38, $p < .001$). Both partner protocol and number of iterations had a statistically significant relation with the effect of punishment on cooperation. For partner protocol, punishment more effectively increased cooperation when participants were reassigned to the same group for several trials than when they were reassigned to a new group each trial (β = .30, p = .001). Punishment was also more effective at increasing cooperation as the number of trials of the dilemma increased (β = .33, $p < .001$). The effect of punishment on cooperation did not vary as a function of group size. Cross-societal variation in trust continued to have a significant positive correlation with the effect size (β = .38, $p < .001$).

Table 3. Correlations Between Cross-Societal Variables Coded for 18 Different Societies

Variables	1	2	3	4	5	6	7
1. Trust	—						
2. GDP	.50*	—					
3. Democracy	.23	.70*	—				
4. Wealth inequality	−.51*	−.27	−.05	—			
5. Market competition	.62*	.70*	.33	−.34	—		
6. Religious attendance	.39	.25	.09	−.44	.57*	—	
7. Norms of cooperation	.37	.55*	.43	.04	.66*	.24	—

Note: All correlations are based on a sample of 18 countries, except for market competition, religious attendance, and norms of cooperation, which include 17 countries.
*$p < .05$.

In Models 3, 4, and 5 we added GDP per capita, wealth inequality, and the democracy index, respectively. In Model 3, GDP per capita, had a significant positive relation with the effect size (β = .21, p = .046). Punishment had a stronger positive effect on cooperation in wealthier countries than it did in less wealthy countries. However, when the democracy index was added to Model 4 then GDP per capita did not have a significant positive relation with the effect size (β = −.03, p = .840). In this model, democracy had a marginally significant positive relation with the effect size (β = .28, p = .057), which indicates that punishment was more effective at promoting cooperation in high-democratic counties than it was in low-democratic countries. In Model 5, the Gini index did not relate to the effect size. In Model 5, trust continued to have a significant positive relation with the effect size (β = .35, p = .006).[5] It is important to note that cross-societal variation in trust had a significant positive relation with the effect size across all five models. This indicates that punishment was relatively more effective at

Table 4. Random Effects Multiple Regression Models Predicting the Effect of Punishment on Cooperation in Public Goods Experiments (d Value)

	Model															
	1		2		3		4		5		6		7		8	
Variable	β	p	β	p	β	p	β	p	β	p	β	p	β	p	β	p
Trust	.33	.001	.38	.001**	.26	.016	.28	.009	.35	.006	.26	.016	.33	.002	.23	.035
Partner protocol			.30	.001	.31	.001	.29	.001	.32	.001	.28	.002	.33	.001**	.28	.002
Iterations			.33	.001**	.28	.003	.28	.003	.29	.002	.27	.004	.27	.004	.26	.005
Group size			.00	.989	−.04	.595	−.04	.609	−.03	.688	−.05	.549	−.01	.946	−.06	.509
GDP					.21	.046	−.03	.840	−.04	.826	−.07	.721	−.16	.372	−.12	.481
Democracy							.28	.057	.29	.056	.27	.064	.38	.013	.20	.188
Wealth inequality									.11	.300						
Market competition											.06	.721				
Religious attendance													.06	.447		
Norms of cooperation															.24	.073
	R^2 = .11**		R^2 = .38**		R^2 = .41**		R^2 = .43**		R^2 = .44**		R^2 = .43**		R^2 = .48**		R^2 = .46**	

Note: For partner protocol, 0 = stranger-matching design, 1 = partner-matching design. For Models 6, 7, and 8, we excluded wealth inequality (Gini index) and replaced it with one additional cross-country coding. This strategy was taken because we had a sample of only 69 effect sizes and wanted to keep no more than 7 predictor variables in any given model.
**$p < .001$.

increasing cooperation in countries with higher levels of trust in others.

In Models 6 through 8, we dropped the Gini index and added market competitiveness, religious attendance, and norms of civic cooperation, respectively. None of these variables had a significant relation with the effect size, although norms of civic cooperation had a marginally significant positive relation with the effect size ($\beta = .24$, $p = .073$). Punishment was more effective at promoting cooperation in countries with stronger norms for cooperation. Trust continued to have a positive relation with the effect size after controlling for these variables. The relation between trust and the effect size for each country is displayed in Figure 1.

Discussion

The primary purpose of the present meta-analysis was to test two competing predictions about how cross-societal differences in trust may moderate the relation between punishment and its effect on cooperation in public goods experiments. Using a meta-analytic approach involving more than 7,000 participants from 18 societies, we provide strong evidence for the perspective that trust and social norm enforcement may reinforce each other in securing and promoting cooperation in large-scale societies. This is indicated by the positive relation between cross-societal trust and the effect of punishment on cooperation in public goods. The broad conclusion to this puzzle of punishment is clear: The effectiveness of punishment in promoting cooperation in a public goods experiment is greater in societies with high trust, rather than low trust.

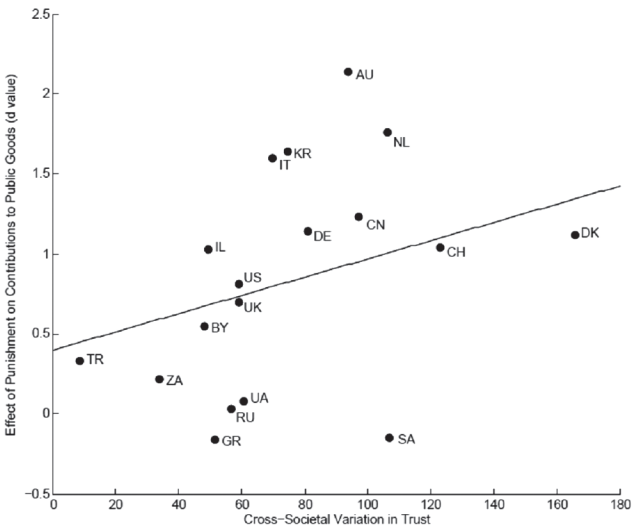

Fig. 1. The scatterplot displays the relation between the effect of punishment on contributions to public goods (*d* value) and cross-societal variation in trust across 18 societies, after controlling for three between-study and three between-country variables (Model 8 reported in Table 2). AU = Australia, BY = Belarus; CN = China; DE = Germany; DK = Denmark; GR = Greece; IL = Israel; IT = Italy; NL = Netherlands; RU = Russia; SA = Saudi Arabia; ZA = South Africa; KR = South Korea; CH = Switzerland; TR = Turkey; UA = Ukraine; UK = United Kingdom; US = United States.

Such evidence is consistent with recent findings that the use of punishment to promote cooperation among strangers is closely linked to a society's social norms for contributions to public goods and the punishment of free-riders (Henrich, Ensminger, et al., 2010; Henrich et al., 2006; Herrmann et al., 2008). Moreover, the present findings unpack the puzzle of punishment even further by providing novel support for the perspective that societal levels of trust and the enforcement of social norms are mutually reinforcing. In the following sections, we address the relation between trust and norm enforcement; discuss some theoretical and practical implications this relation has for understanding the workings of small groups, organizations, and societal institutions; and suggest directions for future research.

Trust, punishment, and cooperation

Earlier, we reasoned that in high-trust societies, societal members may adhere to norms that encourage both cooperation and the punishment of free riders (Coleman, 1988). However, in low-trust societies, these norms may be less strongly shared and enforced, so punishment may be less effective in these societies. We propose that it is unlikely that a single causal direction underlies the relationship between trust and norm enforcement but that both variables may be mutually reinforcing in promoting cooperation. Below, we offer some theoretical perspectives that might help us explain the mechanisms underlying this relation.

One perspective centers on the causal role of trust in facilitating norm enforcement. Trust includes a benevolent view of others' intentions, which is likely to affect judgments of others who engage in the costly punishment of non-cooperators. In high-trust societies, punishment is likely to be viewed as attempts to enforce cooperative norms, which encourages others to behave according to such norms. For example, Balliet and colleagues (2011) found support for their reasoning that punishments were more effective at promoting cooperation when they were more costly (versus free of costs) to administer. From this perspective, it is possible that the primary mechanism is rooted in communication: A willingness to pay a cost to punish others, especially non-cooperative others, is likely to be viewed as a strong concern with collective outcomes. At the same time, such benevolent views of costly punishment may be more likely to occur in societies that contain higher amounts of trust in others, which we conceptualized earlier in terms of beliefs about benevolence toward the self and others.

Thus, trust in others may result in benevolent perceptions of cooperative norm enforcers, which may have two effects that sustain cooperative norms. First, such benevolent perceptions of punishers may result in an enhanced status or reputation of the punishers (dos Santos, Rankin, & Wedekind, 2011; Nelissen, 2008), which may subsequently reinforce the punishment of noncooperators. Second, in high-trust societies, people may be more inclined to perceive their own noncooperation as a violation of a social norm and feel guilty for having violated such a norm (McGraw, 1987). Therefore, individuals may begin to cooperate in response to punishment. However, in low-trust societies that emphasize strong family ties and lack norms of cooperation with strangers, people may respond negatively to being punished for not cooperating with other strangers. Instead of increasing their cooperation they may choose to retaliate and punish others who cooperate (i.e., antisocial punishment; see Gintis, 2008).

Another perspective is that punishment promotes norms of cooperation, which leads to greater amounts of societal trust. That is, although trust may promote effective norm enforcement, so may effective norm enforcement promote trust. Indeed, Ostrom and Ahn (2008) claimed that trust is an outcome of successful collective action, which increases trust in others. So what facilitates successful collective action before the emergence of trust? Prior research on small-scale societies suggests that informal norm enforcement of cooperative norms might be promoted by both the extent of market exchange between strangers and the extent that societal members adopt a major world religion (Henrich, Ensminger, et al., 2010). We found that the extent of market competition or participation in religion (at least as measured by church attendance) did not predict the effectiveness of punishment to promote cooperation.[6] Thus, these cross-societal differences may not necessarily extend to the present sample of countries. Yet, in contrast to prior research on small-scale societies, our analysis included large-scale modern industrialized societies that all possess relatively extensive market economies and exposure to world religions.

Another possibility is that norms and formal institutions that encourage individuals to establish and maintain social relations beyond an extended kin network may promote cooperation among unrelated others and the enforcement of such cooperative group norms (Coleman, 1988; Fukuyama, 1995; Putnam, 1993). Countries that possess norms for maintaining strong family ties may inhibit interactions between unrelated others which may undermine the development of cooperative norms between strangers (Fukuyama, 1995; Putnam, 1993).

There is indeed some recent support for this perspective. A case in point is research by Ermisch and Gambetta (2010) which revealed that people who tend to spend more time with their family and on family care displayed less trust in strangers, compared to individuals who spend less time with their family. These findings raise intriguing questions: For example, how exactly might family ties inhibit trust and cooperation with strangers? And assuming that there are important benefits to forming and maintaining exchange relations with others beyond an extended kin network, an important question is how people come to interact cooperatively with strangers, develop norms for cooperative interaction with strangers, and provide a climate for effective norm enforcement. Future research on such issues may go far in explaining cross-societal variation in cooperation.

Trust and norm enforcement: A social capital perspective

It is interesting to view the present findings in light of the conceptualization and workings of social capital—a widely used concept with broad implications for the functioning and development of a society (see Adler & Kwon, 2002; Coleman, 1988; Fukuyama, 1995; Putnam, 1993). Trust, norms of cooperation, and the effective enforcement of norms are conceptualized as interrelated components of social capital (Coleman, 1988; Ostrom & Ahn, 2008; Portes, 1998), which holds important implications for understanding the functioning of societies, including market economies (Knack & Keefer, 1997) and government (Ostrom & Ahn, 2008; Putnam, 1993).

The present research provides evidence that effective norm enforcement for cooperative behavior, which results in greater success in the provision of public goods, positively relates to a society's level of trust and norms of civic cooperation. Moreover, in additional analyses we combined the measures of trust and social norms to create an index of social capital and found that this had a strong positive relation with the effect size (see supplemental materials). As such, we found novel support for the position that trust and the enforcement of cooperative norms are positively associated, as would be expected from a social capital perspective.

An implication of the present findings is that societies that succeed in establishing cooperative norms and greater trust among strangers should be more capable of expanding social networks embedded in those societies and beyond (Buchan et al., 2009), by providing a formal or informal climate in which new members will behave according to the cooperative group norms that exist in those networks (Coleman, 1988). In this perspective, the prevalence of informal norm enforcement facilitates the expansion of cooperative, social networks, further bolstering the development of social capital.

Norm enforcement and democracy

Cooperation among societal members and the willingness to punish free riders may be essential to the well-being of large-scale institutions (Henrich, Ensminger, et al., 2010; Putnam, 1993). Successful economies and governments may occur as a result of a multitude of societal members engaging in successful collective action in the production of public goods (Ostrom & Ahn, 2008). For example, democratic governments thrive in societies where citizens freely engage in public life by participating in debates, electing representatives, and joining political parties (La Due Lake & Huckfeldt, 1998; Putnam, 1993).

Theory and research suggest that social capital embodied in social networks within a society, such as norms that are sanctioned, are positively related to political engagement and participation, which is the cornerstone of a successful democracy (La Due Lake & Huckfeldt, 1998; Putnam, 1993). Thus, an implication is that successful societal institutions may gain and maintain their success via informal social processes realized among the social networks within those societies. The present findings provide some preliminary evidence in support of this basic argument. Specifically, with those lines of reasoning in mind, we explored the relation between the extent of political participation by societal members and the effect of punishment on cooperation (see supplementary materials). We found that cooperative norm enforcement was more effective in societies with higher levels of political involvement by societal members, even when controlling for societal differences in trust and norms of cooperation.

The present research is promising, in that it clearly displays that the informal enforcement of cooperative norms positively relates to participation in democratic societies. Yet, further research is needed to more closely examine the relation between human cooperation and the functioning of large-scale institutions, such as governments. Doing so will help develop tailored approaches to encouraging cooperation—potentially including such behaviors as political participation by societal members. This issue is all the more pressing given the many societies that are currently making the transition from authoritarian regimes to more democratic states. Strategies that harness the power of informal social processes to support the workings of such institutions may provide a key solution to a long-standing challenge faced by such societies.

Societal implications

Our findings hold important implications for the interactions within small groups, to understanding the successful workings of organizations and larger societal institutions. One important direction for future research is to examine whether our findings readily extend to small groups and organizations, such that those that contain greater amounts of trust also possess an enhanced ability for the effective enforcement of cooperative norms. One implication of our work is that organizations that have successfully developed trust through organizational members participating in informal norm enforcement may well consider an organizational structure that allows employees to monitor and sanction their own performance of specific behaviors, thereby reducing the need to implement costly monitoring and sanctioning systems. In certain organizational contexts where centralized monitoring of team performance is impossible, selecting organizational members based on individual differences in trust may stabilize group norms of performance and ultimately enhance productivity.

An intriguing implication of the present research is that the success of large-scale institutions may be founded on harnessing the stabilizing effects of informal social processes promoting cooperation (Cook, Hardin, & Levi, 2005). Indeed, authorities and institutions may benefit by recognizing some degree of self-determination by local communities (Ostrom, 1990). So in many ways, it may well be the local, informal enforcement of social norms of cooperation that maintains trust in others, which then provides the fertile ground for broader institutional mechanisms to further ensure the trustworthiness of others via increased monitoring and sanctioning (Cook et al., 2005). Thus, cost-effective institutional solutions may take advantage of the knowledge and ability of individuals to solve their own public goods and commons dilemmas (Ostrom, 1990). It may simply be unnecessary for a strong and costly Leviathan to monitor and sanction societal members' behavior in order to sustain cooperation. It is all the more interesting that these local, interpersonal mechanisms are so clearly uncovered in the constraints of a laboratory where strangers are faced with the challenges of a public good dilemma.

Limitations and future directions

There are some limitations worth noting that may be addressed in future research. First, the measure of trust taken by the WVS is not without criticisms. Some definitions of trust presume that people can be both trusting and suspicious of others (Yamagishi, 2011), and there may be systematic differences across cultures in whom people think about when they are asked to think about "most people." Although this question has limitations, responses to this question have been found to predict cross-societal differences in wealth (e.g., Knack & Keefer, 1997) and civic participation (La Porta, Lopez-de-Silanes, Shleifer, & Vishny, 1997). Nonetheless, future work will strengthen our conclusions by examining cultural differences in trust

by using different measures and relating this to the effect of punishment on cooperation in public goods.

Another limitation is that the sample of experimental studies included in the meta-analysis were conducted with university students. Although this may provide a benefit by drawing relatively comparable samples across societies, university students in large-scale modern industrialized societies tend not to capture the broad range of cultural differences observed across large- and small-scale societies (Henrich, Heine, & Norenzayan, 2010). Future research should determine whether our results generalize to the broader range of cultures present in the world.

Finally, although the experimental studies all involved a highly standardized experimental procedure, the experiments were conducted by different experimenters in different labs. Certainly, a gold standard of cross-cultural research is to have a single experimenter develop and apply the same procedure across societies, with a close look at language and for economic experiments such as these, the value of outcomes used in specific studies (see Herrmann et al., 2008). Nonetheless, despite these limitations, a meta-analytic approach, with the ability to look across more studies in more countries than is often the case at the study level, remains a powerful tool to examine variation in studies across societies and test theory of potential cross-cultural differences.

Concluding Remarks

Free riding is one of the most serious threats to public goods and may therefore pose an important risk to the functioning of groups, organizations, and societies. Although punishment has been identified as a powerful solution to the free-rider problem, the effectiveness of punishment as a tool to promote cooperation varies considerably across societies. The intriguing question now is what we have referred to as a puzzle of punishment: Is punishment more effective in low-trust societies or high-trust societies?

The meta-analysis provides a clear answer: Punishment is more effective in promoting cooperation (and reducing free riding) in high-trust societies than in low-trust societies. We have suggested that punishment, as discussed in terms of norm enforcement, and trust might reinforce each other in promoting cooperation—rather than existing in a simple unidirectional causal relation. Such reinforcement of trust and norm enforcement—that trust provides a supporting environment for norm enforcement, and that effective informal norm enforcement promotes trust—helps us understand why certain groups and collectives have higher and lower levels of cooperation.

The present findings underline the importance of trust, punishment, and cooperation as three ingredients that may help us understand why some societies are on their way to stability and growth, while other societies are not. We found evidence of the positive association between societies' levels of trust and the enforcement of cooperative group norms in experimental social dilemmas, which are often conceptualized as two components of social capital that relate to the successful workings of groups, organizations, and even nations. Furthermore, we found support that the informal punishment of cooperative group norms is more effective in societies involving greater participation of societal members in politics, a hallmark of a healthy democracy. Thus, an intriguing implication is that trust and cooperative norm enforcement, as a form of social capital, result in prosperous large-scale societies, and in many ways "make the world go round."

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Notes

1. Past research has examined two distinct forms of sanctioning systems, punishments administered by a centralized authority (e.g., Baldassarri & Grossman, 2011; Mulder, van Dijk, De Cremer, & Wilke, 2006) and punishments delivered by peers (Fehr & Gächter, 2002). In the present research, we focused on the (costly) peer punishment administered in the Fehr and Gächter (2002) research protocol, because a standardized experimental protocol has been developed and used in most studies on this topic. This standardized protocol has also been used in various countries over the past decade, thus allowing for a relatively straightforward empirical analysis to illuminate this puzzle of punishment. Therefore, when we refer to *punishment* throughout this article, we refer to this form of punishment.
2. When we used the WVS, we were able to code all countries included in the meta-analysis for religious participation, norms of civic cooperation, and trust, except Greece and Denmark. For these two countries, we used the EVS, which asks the same questions, to calculate and code these variables.
3. The dataset can be obtained from Daniel Balliet's Web site at http://www.psy.vu.nl/nl/over-de-faculteit/medewerkers-alfabetisch/medewerkers-a-b/d-balliet/index.asp.
4. We examined the possibility that cross-societal variation in trust has a nonlinear relation with the effect of punishment on cooperation. We examined both the quadratic and logarithmic transformations of trust and their relations with the effect size and did not find any support for a nonlinear relation.
5. We also computed the interaction between trust and democracy and added the interaction term to the model predicting the effect size. The interaction was not a significant predictor of the effect size.
6. Additional research has found that religious beliefs relate to higher amounts of cooperation and wealth across societies (Atkinson & Bourrat, 2011; Barro & McCleary, 2003). In

the supplementary material, we coded and analyzed whether cross-societal variation in religious beliefs predicted the effect of punishment on cooperation. Religious beliefs did not have a significant relation with the effect size.

References

*Indicates study was used in the meta-analysis

Adler, P. S., & Kwon, S.-W. (2002). Social capital: Prospects for a new concept. *Academy of Management Review*, *27*, 17–40.

Atkinson, Q. D., & Bourrat, P. (2011). Beliefs about God, the afterlife and morality support the role supernatural policing in human cooperation. *Evolution & Human Behavior*, *32*, 41–49.

Baldassarri, D., & Grossman, G. (2011). Centralized sanctioning and legitimate authority promote cooperation in humans. *Proceedings of the National Academy of Sciences, USA*, *108*, 11023–11027.

Balliet, D. (2010). Communication and cooperation in social dilemmas: A meta-analytic review. *Journal of Conflict Resolution*, *54*, 39–57.

Balliet, D., Mulder, L. B., & Van Lange, P. A. M. (2011). Reward, punishment, and cooperation: A meta-analysis. *Psychological Bulletin*, *137*, 594–615.

Balliet, D., & Van Lange, P. A. M. (2012, December 10). Trust, conflict, and cooperation: A meta-analysis. *Psychological Bulletin*. Advance online publication. doi:10.1037//a003 0939

Barro, R. J., & McCleary, R. M. (2003). Religion and economic growth across countries. *American Sociological Review*, *68*, 760–781.

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

Bond, M. H., Leung, K., Au, A., Tong, K.-K., de Carrasquel, S. R., Murakami, F., . . . Rees Lewis, J. (2004). Culture-level dimensions of social axioms and their correlates across 41 cultures. *Journal of Cross-Cultural Psychology*, *35*, 548–570.

*Bornstein, G., & Weisel, O. (2010a). Punishment, cooperation, and cheater detection in "noisy" social exchange. *Games, 1*, 18–33. doi:10.3390/g1010018

*Bornstein, G., & Weisel, O. (2010b). [A replication of the study reported in Punishment, cooperation, and cheater detection in "noisy" social exchange]. Raw Data, The Hebrew University of Jerusalem, Israel.

Buchan, N., Grimalda, G., Wilson, R., Brewer, M., Fatas, E., & Foddy, M. (2009). Globalization and human cooperation. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 4138–4142.

*Carpenter, J. P. (2007). Punishing free-riders: How group size affects mutual monitoring and the provision of public goods. *Games and Economic Behavior, 60*, 31–51. doi:10.1016/j.geb.2006.08.011

*Carpenter, J., & Matthews, P. H. (2009). What norms trigger punishment? *Experimental Economics, 12*, 272–288. doi:10.1007/s10683-009-9214-z

*Carpenter, J., Matthews, P., & Ong'ong'a, O. (2004). Why punish? Social reciprocity and the enforcement of prosocial norms. *Journal of Evolutionary Economics, 14*, 407–429. doi:10.1007/s00191-004-0212-1

*Casari, M., & Luini, L. (2009). Cooperation under alternative punishment institutions: An experiment. *Journal of Economic Behavior & Organization, 71*, 273–282. doi:10.1016/j.jebo.2009.03.022

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics, 14*, 47–83. doi:10.1007/s10683-0109257-1

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, *94*, 95–120.

Cook, K. S., Hardin, R., & Levi, M. (2005). *Cooperation without trust?* New York, NY: Russell Sage Foundation.

Cooper, H. M. (1998). *Synthesizing research: A guide for literature reviews* (3rd ed.). Thousand Oaks, CA: Sage.

Dirks, K. T. (1999). The effects of interpersonal trust on work group performance. *Journal of Applied Psychology*, *84*, 445–455.

Doney, P. M., Cannon, J. P., & Mullen, M. R. (1998). Understanding the influence of national culture on the development of trust. *Academy of Management Review*, *23*, 601–620.

dos Santos, M., Rankin, D. J., & Wedekind, C. (2011). The evolution of punishment through reputation. *Proceedings of the Royal Society Biological Sciences*, *278*, 371–377. doi:10.1098/rspb.2010.1275

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455–463.

Economist Intelligence Unit. (2010). *Democracy index 2010*. London, England: Author. Retrieved from http://graphics.eiu.com/PDF/Democracy_Index_2010_web.pdf

*Egas, M., & Riedl, A. (2008). The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society Biological Sciences*, *275*, 871–878. doi:10.1098/rspb.2007.1558

Ermisch, J., & Gambetta, D. (2010). Do strong family ties inhibit trust? *Journal of Economic Behavior & Organization*, *75*, 365–372.

*Ertan, A., Page, T., & Putterman, L. (2009). Who to punish? Individual decisions and majority rule in mitigating the free rider problem. *European Economic Review*, *53*, 495–511. doi:10.1016/j.euroecorev.2008.09.007

*Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, *90*, 980–994.

*Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137–140.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economic Letters*, *71*, 397–404.

Fukuyama, F. (1995). *Trust*. New York, NY: The Free Press.

*Fuster, A., & Meier, S. (2009). Another hidden cost of incentives: The detrimental effect on norm enforcement. *Management Science*, *56*, 57–70. doi:10.1287/mnsc.1090.1081

*Gächter, S., & Herrmann, B. (2009). Reciprocity, culture and human cooperation: Previous insights and a new cross-cultural experiment. *Philosophical Transactions of the Royal Society Biological Sciences*, *364*, 791–806. doi:10.1098/rstb.2008.0275

*Gächter, S., & Herrmann, B. (2011). The limits of self-governance when cooperators get punished: Experimental evidence from urban and rural Russia. *European Economic Review*, *55*(2), 193–210. doi:10.1016/j.euroecorev.2010.04.003

Gächter, S., Herrmann, B., & Thöni, C. (2004). Trust, voluntary cooperation, and socio-economic background: Survey and experimental evidence. *Journal of Economic Behavior & Organization*, *55*, 505–531. doi:10.1016/j.jebo.2003.11.006

*Gächter, S., Renner, E., & Sefton, M. (2008). The long-run benefits of punishment. *Science*, *322*, 1510. doi:10.1126/science.1164744

*Gächter, S., & Thoni, C. (2005). Social learning and voluntary cooperation among like minded people. *Journal of European Economic Association*, *3*, 303–314. doi:10.1162/jeea.2005.3.2-3.303

Gintis, H. (2008). Punishment and cooperation. *Science*, *319*, 1345–1346.

Gürerk, Ö., Irlenbusch, B., & Rockenbach, B. (2006). The competitive advantage of sanctioning institutions. *Science*, *312*, 108–111.

Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., . . . Ziker, J. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science*, *327*, 1480–1484.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–83.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., . . . Ziker, J. (2006). Costly punishment across human societies. *Science*, *312*, 1767–1770.

*Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science, 319*, 1362–1367. doi:10.1126/science.1153808

Inglehart, R., Basanez, M., & Moreno, A. (1998). *Human values and beliefs: A cross-cultural sourcebook*. Ann Arbor: University of Michigan Press.

Knack, S. (2001). *Trust, associational life and economic performance*. Unpublished manuscript, World Bank, Washington, DC.

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

*Kocher, M. G., Martinsoon, P., & Visser, M. (2008). Does stake size matter for cooperation and punishment? *Economic Letters*, *99*, 508–511. doi:10.1016/j.econlet.2007.09.048

*Kroll, S., Cherry, T. L., & Shogren, J. F. (2007). Voting, punishment, and public goods. *Economic Inquiry*, *45*, 557–570. doi:10.1111/j.1465-7295.2007.00028.x

La Due Lake, R., & Huckfeldt, R. (1998). Social capital, social networks, and political participation. *Political Psychology*, *19*, 567–584.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. W. (1997). Trust in large organizations. *The American Economic Review*, *87*, 333–338.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. London, England: SAGE.

Marlowe, F. W., Berbesque, J. C., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., . . . Tracer, D. (2008). More "altruistic" punishment in larger societies. *Proceedings of the Royal Society B: Biological Sciences*, *275*, 587–590.

McGraw, K. M. (1987). Guilt following transgression: An attribution of responsibility approach. *Journal of Personality and Social Psychology*, *53*, 247–256.

Mulder, L. B., van Dijk, E., De Cremer, D., & Wilke, H. A. M. (2006). Undermining trust and cooperation: The paradox of sanctioning systems in social dilemmas. *Journal of Experimental Social Psychology*, *42*, 147–162. doi:10.1016/j.jesp.2005.03.002

*Myers, C. D. (2009). *Participation and punishment*. Unpublished manuscript, Princeton University, Princeton, NJ.

Nelissen, R. M. A. (2008). The price you pay: Cost-dependent reputation effects of altruistic punishment. *Evolution & Human Behavior*, *29*, 242–248. doi:10.1016/j.evolhumbehav.2008.01.001

*Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, *92*, 91–112. doi:10.1016/j.jpubeco.2007.04.008

*Nikiforakis, N., & Normann, H. (2008). A comparative statistical analysis of punishment in public-good experiments. *Experimental Economics*, *11*, 358–369. doi:10.1007/s10683-007-9171-3

*Nikiforakis, N., Normann, H., & Wallace, B. (2010). Asymmetric enforcement of cooperation in a social dilemma. *Southern Economic Journal*, *76*, 638–659. doi:10.4284/sej.2010.76.3.638

*O'Gorman, R., Henrich, J., & Van Vugt, M. (2008). Constraining free riding in public goods games: Designated solitary punishers can sustain human cooperation. *Proceedings of the Royal Society B: Biological Sciences*, *276*, 1–7. doi:10.1098/rspb.2008.1082

*O'Gorman, R., & Van Vugt, M. (2010). [Does allowing participants the option to choose their punisher increase contributions in a public goods game]. Raw Data, University of Essex, Colchester, England.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. New York, NY: Cambridge University Press.

Ostrom, E., & Ahn, T. K. (2008). The meaning of social capital and its link to collective action. In D. Castiglione, J. W. van Deth, & G. Wolleb (Eds.), *Handbook on social capital* (pp. 17–35). Northampton, MA: Edward Elgar.

Ostrom, E., Walker, J., & Gardner, R. (1992). Covenants with and without the sword: Self-governance is possible. *American Political Science Review*, *86*, 404–417.

*Page, T., Putterman, L., & Unel, B. (2005). Voluntary association in public goods experiments: Reciprocity, mimicry, and efficiency. *The Economic Journal*, *115*, 1032–1053. doi:10.1111/j.1468-0297.2005.01031.x

*Patel, A., Cartwright, E., & Van Vugt, M. (2010). *Punishment cannot sustain cooperation in a public good game with free-rider anonymity*. Unpublished Manuscript, University of Gothenburg, Gothenburg, Sweden.

Portes, A. (1998). Social capital: Its origins and applications in modern sociology. *Annual Review of Sociology*, *24*, 1–24.

Putnam, R. (1993). *Making democracy work*. Princeton, NJ: Princeton University Press.

*Rand, D. G., Dreber, A., Ellingsen, T., Fudenberg, D., & Nowak, M. A. (2009). Positive interactions promote public cooperation. *Science*, *325*, 1272–1275.

*Reuben, E., & Riedl, A. (2009). *Enforcement of contribution norms in public goods games with heterogeneous populations*. Unpublished Manuscript, Columbia University, New York, NY.

Rockenbach, B., & Milinski, M. (2006). The efficient interaction of indirect reciprocity and costly punishment. *Nature*, *444*, 718–723.

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, *35*, 651–665.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*, 393–404.

Schwab, K. (2012). *The global competitiveness report, 2011–2012*. Geneva, Switzerland: World Economic Forum. Retrieved from http://www3.weforum.org/docs/WEF_GCR_Report_2011-12.pdf

*Sefton, M., Shupp, R., & Walker, J. M. (2007). The effect of rewards and sanctions in provision of public goods. *Economic Inquiry*, *45*, 671–690.

*Tan, F. (2008). Punishment in a linear public good game with productivity heterogeneity. *De Economist, 156*, 269–293. doi:10.1007/s10645-008-9094-1O

United Nations Development Programme. (2010). *Human development report 2010*. Houndmills, England: Palgrave Macmillan. Retrieved from http://hdr.undp.org/en/media/HDR_2010_EN_Complete_reprint.pdf

Van Lange, P. A. M., Balliet, D., Parks, C. D., & Van Vugt, M. (in press). *Social dilemmas: The psychology of human cooperation*. New York, NY: Oxford University Press.

Yamagishi, T. (1986). The provisioning of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, *51*, 110–116.

Yamagishi, T. (1988). The provision of a sanctioning system in the United States and Japan. *Social Psychology Quarterly*, *51*, 265–271.

Yamagishi, T. (2011). *Trust*. Tokyo, Japan: Springer.