Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Habits of Virtue: Creating Norms of Cooperation and Defection in the Laboratory

Alexander Peysakhovich, David G. Rand

To cite this article:
Alexander Peysakhovich, David G. Rand (2016) Habits of Virtue: Creating Norms of Cooperation and Defection in the Laboratory. Management Science 62(3):631-647. http://dx.doi.org/10.1287/mnsc.2015.2168

MANAGEMENT SCIENCE
Vol. 62, No. 3, March 2016, pp. 631–647
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.2015.2168

Habits of Virtue: Creating Norms of Cooperation and Defection in the Laboratory

Alexander Peysakhovich
Program for Evolutionary Dynamics, Harvard University, Cambridge, Massachusetts 02138; and
Department of Psychology, Yale University, New Haven, Connecticut 06520, alex.peys@gmail.com

David G. Rand
Department of Psychology, Department of Economics, School of Management, Yale University,
New Haven, Connecticut 06520, david.rand@yale.edu

What explains variability in norms of cooperation across organizations and cultures? One answer comes from the tendency of individuals to internalize typically successful behaviors as norms. Different institutional structures can cause different behavioral norms to be internalized. These norms are then carried over into atypical situations beyond the reach of the institution. Here, we experimentally demonstrate such spillovers. First, we immerse subjects in environments that do or do not support cooperation using repeated prisoner's dilemmas. Afterwards, we measure their intrinsic prosociality in one-shot games. Subjects from environments that support cooperation are more prosocial, more likely to punish selfishness, and more trusting in general. Furthermore, these effects are most pronounced among subjects who use heuristics, suggesting that intuitive processes play a key role in the spillovers we observe. Our findings help to explain variation in one-shot anonymous cooperation, linking this intrinsically motivated prosociality to the externally imposed institutional rules experienced in other settings.

Data, as supplemental material, are available at http://dx.doi.org/10.1287/mnsc.2015.2168.

Keywords: economics: behavior and behavioral decision making; economics: game theory and bargaining theory; organizational studies: decision making; games–group decisions
History: Received April 5, 2014; accepted January 20, 2015, by Uri Gneezy, behavioral economics. Published online in *Articles in Advance* September 9, 2015.

1. Introduction

The tendency to cooperate, or to pay a personal cost to give a benefit to, is not a constant between groups: norms of cooperation and trust vary markedly across organizations (Leana and Van Buren 1999, McAllister 1995, Rousseau et al. 1998), countries (Cappelen et al. 2013, Ellingsen et al. 2012, Henrich et al. 2010, Herrmann et al. 2008), and cultures (Gächter et al. 2010, Henrich et al. 2005, Sapienza et al. 2006). Here, we use economic game experiments to shed light on the origins of this heterogeneity. We argue that cooperative norms (i.e., individual-level conceptions of appropriate behavior)[1] and expectations about the behavior of others are driven, at least in part, by spillovers from representative daily life interactions.[2] Individuals who primarily interact in environments where the "rules of the game" make cooperation advantageous get in the habit of cooperating and, as a result, are more cooperative even in one-shot interactions (such as those in most laboratory games). By this logic, institutional differences across groups lead to differences in cooperative norms between those

[1] There are many different uses of the word "norm" in different literatures. Here, we use *norm* to refer to an individual's internalized conception of what behavior is appropriate in a given setting or range of settings (manifested in terms of that individual's preferences, in addition to beliefs). By this definition, holding a particular norm makes one feel personally compelled to engage in behaviors that the norm deems to be appropriate, and it also makes one upset when other people do not follow the norm and instead engage in behaviors seen as inappropriate (Fehr and Fischbacher 2004a, b; Jordan et al. 2014). For example, individuals with a norm prescribing cooperation will be inclined to both cooperate and punish noncooperators, even at a cost. Thus these norms form the underpinnings of social preferences.

[2] Spillover effects occur when subjects extrapolate from experience in one domain to guide behavior in other domains with different incentive structures (e.g., Rand et al. 2014b). This includes generalization that occurs outside the lab, where strategies from typical settings with future consequences are applied to atypical situations where risk-free exploitation is possible, as well as generalization that occurs inside the lab, either from typical settings outside the lab to one-shot anonymous lab games or from repeated games in the lab to subsequent one-shot games in the lab (as in our experiments).

groups:[3] members of groups with institutions that successfully incentivize cooperation most of the time continue to behave more cooperatively even in interactions where no such institutional incentives exist.

In this paper we provide empirical evidence for this line of reasoning with a set of experiments directly demonstrating how randomly assigned institutions can change norms and create "cultural differences" in the lab. Each experiment has the same basic structure: first, subjects are assigned to interact in an environment where, because of the rules of interaction, cooperative equilibria either are strongly supported or do not exist (Stage A). After experience in this environment, subjects take part in a battery of one-shot anonymous economic games used to study cooperation (Stage B). We then look for the impact of random assignment to experimental environment in Stage A on behavior in Stage B.

To create Stage A lab environments that favor cooperation or noncooperation, we use infinitely repeated prisoner's dilemma games. The extent to which an environment supports cooperation can be manipulated using many mechanisms other than repeated interactions, including reputation, sanctions, partner choice, intergroup competition, and formal institutions (for a review, see Jordan et al. 2015). We chose repeated games as a model of the future consequences created by these mechanisms because the determinants of the emergence (and stability) of cooperation in repeated games are well understood both experimentally (Dal Bó 2005, Dal Bó and Fréchette 2011, Fudenberg et al. 2012, Rand and Nowak 2013) and theoretically (Blonski et al. 2011; Fudenberg and Maskin 1986, 1990; Mailath and Samuelson 2006). This allows us to select combinations of payoffs and continuation probabilities that lead to high levels of cooperation in one treatment (the "C-Culture" treatment) and low levels of cooperation in the other (the "D-Culture" treatment).

In Experiment 1, the repeated prisoner's dilemmas of Stage A are followed by a Stage B consisting of one-shot anonymous cooperation games: the public goods game, trust game, dictator game, and ultimatum game. Each of these games involves one or more choices of whether to transfer money from oneself to one or more others (i.e., to act prosocially). We show that subjects randomized into the C culture in Stage A are substantially more prosocial in the Stage B games compared with subjects randomized into the D-culture treatment. Importantly, this is true even in the dictator game (DG), where the recipient is passive and therefore expectations about the decisions of others play no role. We also provide several pieces of evidence that expectations about the *type* (i.e., level of altruism) of one's Stage B coplayer do not explain our treatment effect. One such piece of evidence comes from replicating our treatment's effect on the DG in a supplemental experiment using an online, nonstudent subject pool where Stage B recipients are complete strangers, do not participate in Stage A, and take no actions in the experiment whatsoever (other than receiving any money given to them). This replication shows that the treatment effect cannot be explained by reciprocity toward the individuals one interacted with in Stage A, and it speaks against inferring that one's DG recipient would have been more or less altruistic based on the observed level of cooperation in one's Stage A partners. Thus, in these experiments, we demonstrate that the rules of the game in one setting can strongly influence behavior in other unrelated settings.

We also explore the cognitive mechanism through which these spillovers occur. To do so, we take a dual-process perspective and conceptualize decision making as the result of interactions between intuitive and deliberative processes (Epstein et al. 1996, Gilovich et al. 2002, Kahneman 2003, Tversky and Kahneman 1974). Intuitive processes are fast, automatic, emotional, and heuristic in nature: intuitions often favor behaviors that are advantageous in typical settings (Gigerenzer and Goldstein 1996, Gigerenzer et al. 1999). Deliberation, by contrast, is slow, controlled, rational, and tailored to the specific decision under consideration. Thus in *atypical* settings, deliberation can override suboptimal intuitive, heuristic responses (which, although they may be optimal in general, are ill-matched to the atypical setting at hand).

In previous work, this dual-process lens has been applied to cooperation via the social heuristics hypothesis (SHH) (Rand et al. 2014b), which takes theories of cultural evolution and norm internalization (Bowles and Gintis 2002, 2003; Boyd and Richerson 2009; Chudek and Henrich 2011; Gintis 2003; Henrich et al. 2006; Richerson and Boyd 2005) and makes them explicitly dual process. The SHH contends that the internalization of norms occurs via the channel of intuitive processing. By this account, social behaviors that are successful in the course of one's daily life (in the developed world, this is typically cooperation) become internalized as default heuristics. In one-shot anonymous interactions (for example, in the lab), deliberation then leads one to realize that selfishness is actually optimal. Empirical evidence for this theory comes from the finding that experimentally inducing heuristic processing (via time pressure, conceptual priming, or cognitive load) can increase

[3] We use a broad definition of *institution* that encompasses any factor external to the individual that affects incentives (in contrast to norms, which are internalized conceptions). By this definition, repeated interactions, reputation systems, and the threat of punishment by third parties or organizations are all forms of institutions.

prosociality in economic games (Cone and Rand 2014; Cornelissen et al. 2011; Lotz 2015; Rand et al. 2012, 2015b, 2014b; Roch et al. 2000; Schulz et al. 2014) (although other studies have found null effects; see, e.g., Hauge et al. 2015, Tinghög et al. 2013, Verkoeijen and Bouwmeester 2014).

In the context of the present experiments, therefore, the SHH predicts that experiences in Stage A affect behavior in Stage B by remodeling subjects' default responses. Stage A should have less of an effect on Stage B behavior among subjects that are better at overriding their intuitive heuristic-based responses. Consistent with this prediction, we find that subjects in Experiment 1 who engage in more deliberative thinking (as measured by the "cognitive reflection test"; Frederick 2005) are much less influenced by Stage A when making their Stage B decisions.

In Experiment 2, we ask whether exposure to a cooperative environment in Stage A also affects the willingness to *enforce* cooperation. To investigate this issue, we follow Stage A with a Stage B that consists of punishment games where subjects can pay to reduce the earnings of others (Fehr and Fischbacher 2004b, Fehr and Gächter 2000, Ostrom et al. 1992, Ouss and Peysakhovich 2015). We find that subjects randomized into the C-Culture in Stage A are more likely to engage in third-party punishment of selfishness than those randomized into the D-Culture; and find some evidence that subjects in the D-Culture may be more likely to engage in anti-social punishment of cooperators in a public goods game. We also replicate the result from Experiment 1 that, as predicted by the SHH, Stage A has a much smaller effect on Stage B behavior among subjects who engage in more deliberative thinking. Thus, in Experiment 2, we provide evidence that Stage A affects *norms* of cooperation, altering enforcement behavior as well as cooperative choice.

Finally, we ask whether our Stage A manipulation has effects that generalize beyond behavior in economic games. To do so, we have subjects answer a commonly used question regarding generalized trust from the World Values Survey, which has been shown to reflect prosocial orientation (not just beliefs about the trustworthiness of others). Consistent with our Stage B game results, we find that subjects randomized into the D-Culture treatment in Stage A report being significantly less generally trusting of others than those from the C-Culture, and that this effect is more pronounced among subjects who rely on heuristics.

Thus, by exposing our subjects to environments that favor cooperation or defection, we replicate three major results in cross-cultural studies within a single subject pool: variation in prosociality (Henrich et al. 2005, 2010), variation in norm-enforcement (Ellingsen et al. 2012; Gächter et al. 2010; Henrich et al. 2005, 2010, 2006; Herrmann et al. 2008), and variation in generalized trust (La Porta et al. 2001, Putnam 2000). In doing so, we provide causal evidence in support of previous cross-cultural correlational studies suggesting that norms related to cooperation (revealed by play in one-shot economic games) positively covary across cultures with measures of institutional quality such as rule of law (Gächter et al. 2010, Herrmann et al. 2008) and market integration (Henrich et al. 2010):[4] we show that a qualitatively equivalent pattern of results can be generated through random assignment in the laboratory, operating through the channel of heuristic decision making. Thus our experiments support the argument that institutional differences across organizations and cultures can affect internalized norms of prosociality, and demonstrate how norms can be changed through a top-down process driven by institution designers.

The rest of our paper proceeds as follows. In §2, we describe the design of Stage A, which is the same in both Experiments 1 and 2. In §3, we describe the design of Experiment 1's Stage B, which investigates prosociality, and we present the results of the experiment. In §4, we describe the design of Experiment 2's Stage B, which investigates punishment behavior, and we present the results of the experiment. In §5, we aggregate across studies and investigate treatment effects on generalized trust (rather than game play). In §6, we present a concluding discussion.

2. General Experimental Design for Creating Cultures of Cooperation or Defection (Stage A)

Our two-stage experiments are designed to evaluate the effect of exposure to environments that incentivize cooperation or noncooperation in Stage A on subsequent behavior in one-shot anonymous interactions in Stage B. In Stage A, subjects play a series of stochastically repeated prisoner's dilemma (RPD) games with different partners. At the beginning of each game, subjects are matched in pairs. Each game consists of a random number of rounds. In each round, subjects play a simultaneous PD stage game: both players choose an action, C or D (labeled A and B for the subjects). Subjects are then informed of the decision of their partner and the resulting earnings of each player for the round. They then play another round with the same partner with probability δ; with probability $1-\delta$, the game ends and subjects are rematched

[4] Bowles (1998) surveys more evidence to this effect and presents a model in which institutional factors can affect the endogenous evolution of preferences.

with a new partner for a new game (and informed of this rematching).

To manipulate the extent to which the Stage A environment supports cooperation versus noncooperation, subjects are randomly assigned to one of two treatments: the C-Culture treatment or the D-Culture treatment. The treatments differ both in continuation probabilities, with $\delta = 7/8$ (expected game length of 8 rounds) in the C-Culture and $\delta = 1/8$ (expected game length of 1.14 rounds) in the D-Culture, and the PD game payoffs, with the D-Culture involving a higher temptation payoff from defection exploiting cooperation. The stage game payoffs (shown in monetary units, or MU) are shown in the table below.

C-Culture	C	D	D-Culture	C	D
C	4, 4	0, 5	C	4, 4	0, 6
D	5, 0	1, 1	D	6, 0	1, 1

At the end of the experiment, MU are converted to cash at an exchange rate of \$1 = 30 MU.

An important determinant of whether cooperation emerges in infinitely repeated games is whether the strategy tit-for-tat (TfT) risk dominates always defect (AllD) (Blonski et al. 2011, Rand and Nowak 2013). For our C-Culture, we therefore choose parameters such that TfT strongly risk dominates AllD, and so we expect subjects in the C-culture to learn to cooperate and maintain cooperation in the long run.[5] In the D-Culture treatment, conversely, we choose a specification such that cooperation is not an equilibrium, and therefore we expect subjects to learn to defect.[6] Thus Stage A creates environments in which players consistently cooperate (C-Culture) or consistently defect (D-Culture).

To control for random variation in lengths between sessions, we follow the procedure of Dreber et al. (2008), Fudenberg et al. (2012), and Rand et al. (2015a). For each of the two treatments, we generate a single set of game lengths using the appropriate distribution, and then we use this same set of game lengths in every session. The specific game lengths used are shown in Online Appendix Tables A18 and A19 (available as supplemental material at http://dx.doi.org/10.1287/mnsc.2015.2168). In Experiment 1, subjects play a total of 53 rounds (split into 10 games) in the C-Culture treatment and 51 rounds (split into 45 games) in the D-Culture treatment. In Experiment 2, the Stage B decisions take longer to complete, and so we shorten the total length of Stage A to be approximately 40 PD rounds in total (split into 7 games in the C-Culture and 35 in the D-Culture).

Our goal is to examine the consequences of the resulting habituation and reinforcement of either cooperation or defection on subsequent behavior. A potential confound, however, exists in the form of income effects: if subjects in the C-Culture treatment cooperate much more than those in the D-Culture treatment, Stage A earnings will be higher in the C-Culture treatment. To ensure that any differences we observe in Stage B result from acculturation to cooperation or defection rather than any consequences of differences in income, we vary the size of the initial endowment subjects receive at the beginning of Stage A: in the D-Culture treatment, subjects begin with an endowment of 150 MU, whereas in the C-Culture treatment, subjects begin with only 40 MU (50 MU in Experiment 2 to control for shorter Stage A length). As a result, subjects in each condition finish Stage A with similar earnings on average.[7] Thus, our results cannot be explained by players in the C-Culture earning more than players in the D-Culture.

After completing Stage A, subjects enter Stage B and play a battery of one-shot anonymous games commonly used to assess prosociality (Experiment 1) or punishment (Experiment 2). Subjects are given no information about the history of Stage A play of their Stage B interaction partners. The Stage B games are described in more detail in the corresponding sections for each experiment below.

Unless otherwise noted, all analyses reported use linear regression. To correct for within-session correlation induced by Stage A behavior, we cluster standard errors at the session level. To address potential small sample bias resulting from the low number of sessions, we calculate significance levels for each regression by bootstrapping, with 1,000 iterations per bootstrap. This clustering also addresses correlation across decisions made by a particular subject, and replacing session-level clustering with subject-level clustering calculated via standard normal approximations does not qualitatively change any of our results.

[5] The payoff of TfT against a 50–50 mix between Tft and AllD is given by $0.5(4)\cdot(1/(1-\delta))+0.5(0+1\cdot\delta/(1-\delta))=19.5$, whereas the payoff of AllD against the same mix is given by $0.5(5)+0.5(1)+1\cdot(\delta/(1-\delta))=10$. Thus TfT strongly risk dominates AllD.

[6] No cooperative equilibria exist as even in the presence of the harshest possible punishment (grim trigger), the present gain from defecting (2 for sure) outweighs the expected future losses from loss of cooperation $(3\cdot(\delta/(1-\delta))\sim 0.33)$.

[7] Subjects in the C-Culture earned substantially more MU during the RPD than subjects in the D-Culture (Experiment 1: C-Culture, 166.18 MU versus D-Culture, 80.65 MU; Experiment 2: C-Culture, 110.25 MU versus D-Culture, 61.68 MU). However, the difference in initial endowments more than made up for this difference. Thus, upon leaving Stage A, subjects in the C-Culture had actually earned less than subjects in the D-Culture (Experiment 1: C-Culture, 206.18 MU versus D-Culture, 230.65 MU, rank-sum $p = 0.001$; Experiment 2: C-Culture, 160.25 versus D-Culture, 211.68, rank-sum $p < 0.001$).

Table 1 Descriptive Statistics for Each Experiment

	No. of sessions	No. of subjects	Mean age	Female (%)
Experiment 1	6 (3 C-Culture)	96 (44 C-Culture)	21.7	37
Experiment 2	10 (6 C-Culture)	122 (66 C-Culture)	21.8	45

Note. No subjects dropped out of the experiment or are excluded from analysis.

Our experiments were run in the Harvard Decision Sciences Laboratory between April and September 2012 and were implemented using the z-Tree software (Fischbacher 2007). We aimed for approximately 50 subjects per condition, per the recommendations of Simmons et al. (2013). Treatment was randomly assigned at the session level, and subjects who participated in Experiment 1 could not participate in Experiment 2. See Table 1 for descriptive statistics for each experiment.

3. Experiment 1: Effects of Cooperative Environment on Prosociality

3.1. Experimental Design

Experiment 1 tests our prediction that exposure to a laboratory environment where cooperation is a strongly supported equilibrium will lead to more prosociality in subsequent one-shot anonymous games compared with a laboratory environment where cooperation is not an equilibrium. To do so, the RPDs of Stage A described above are followed by a Stage B consisting of four games that are widely used for measuring prosociality (for an overview, see Camerer and Fehr 2002): a public goods game (PGG), a trust game (TG), a dictator game (DG), and an ultimatum game (UG), played in the listed order. We employ the "strategy method": subjects enter decisions for each possible player role of each game; then at the end of Stage B, one game is selected at random to actually be played for money, and subjects are randomly assigned to player roles in that game with their action being determined by the corresponding choice they indicated earlier. Thus subjects receive no feedback about outcomes between different decisions in Stage B, yet their decisions are still incentivized. Subjects are fully informed about this procedure, as well as the fact that earnings will be translated into cash at the same exchange rate used in Stage A.

In the PGG, players interact in groups of four. Each player is given an endowment of 100 MU and chooses how much of it to keep for him- or herself and how much to transfer (contribute) to a common project. All MU transferred are multiplied by an efficiency factor of 1.6 and then divided equally among all four group members.

In the TG, players interact in pairs: a trustor and a trustee, both of whom begin the game with an endowment of 50 MU. The trustor chooses whether or not to transfer her 50 MU to the trustee (binary choice). If the transfer is made, the 50 MU are tripled and given to the trustee. The trustee chooses how many MU to transfer back to the trustor (any amount from 0 to 150 MU) should the trustor make the transfer. In addition to measuring behavior, we also measure trustor expectations in this game: players are asked to provide a guess (unincentivized, per Peysakhovich and Plagborg-Møller 2012) of the average number of MU transferred back by trustees.

In the DG, subjects again are assigned to pairs: a dictator who begins with 100 MU and a recipient who begins with 0 MU. The dictator then chooses how much of her endowment to transfer to the recipient, who is passive and takes no action. Thus there is no role for the dictator's expectations about the behavior of the recipient, and the DG offers a pure measure of prosocial preferences.

In the UG, subjects once again interact in pairs: a proposer and a responder are given 100 MU to divide between them. The proposer chooses how many MU to offer to the responder. The responder indicates the minimum offer she is willing to accept (minimum acceptable offer, or MAO). If the proposer's offer is greater than or equal to the responder's MAO, then the transfer occurs and each player is paid accordingly. If, on the other hand, the proposer's offer is less than the responder's MAO, the offer is rejected and neither player earns anything.

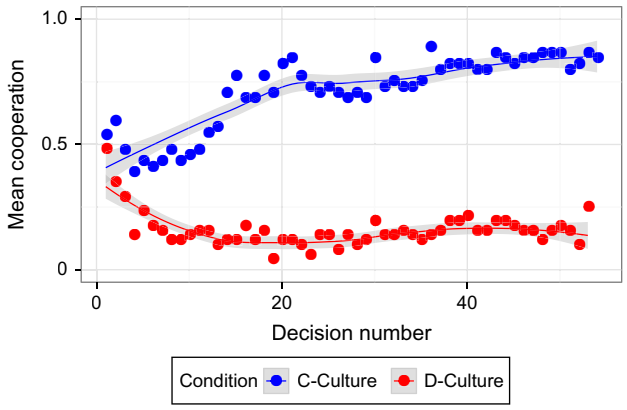
Thus subjects make six decisions in Stage B. Five of these decisions (all but the UG's responder decision) involve transferring money from oneself to one or more others. These five transfer decisions therefore form our Stage B measures of prosociality in one-shot anonymous settings, which we analyze jointly. To compare transfers across different games and roles, we normalize each decision such that the maximum transfer has a value of 1. Thus a normalized value of 1 is assigned to transferring 100 in the PGG, choosing to transfer as the TG trustor, transferring back 150 as the TG trustee, transferring 100 in the DG, and offering to transfer 100 as the UG proposer. Our choice to take these five decisions as measures of prosociality—and not to include the UG MAO—is motivated by a principal component analysis of Stage B play in Experiment 1, which suggests that the five prosociality decisions track together while the UG MAO does not (see the online appendix for details). Further support comes from evidence that these transfer decisions are strongly correlated within an individual but are not correlated with that individual's UG MAO (Peysakhovich et al. 2014, Yamagishi et al. 2012).

To assess the role heuristics play in any effect Stage A might have on Stage B, we have subjects complete the cognitive reflection test (CRT) after finishing Stage B. The CRT is a set of three simple math problems with intuitively compelling but incorrect answers (Frederick 2005).[8] We use CRT scores as a proxy for propensity to engage in intuitive thinking (and thus to follow one's heuristic response) versus stopping to think and thus potentially overriding one's heuristic response.[9] Finally, to assess whether any treatment effect is driven by changing subjects' general affect or mood, subjects in three sessions ($N = 48$, 24 in two C-Culture sessions and 24 in one D-Culture session) are asked, "How would you describe your mood right now?" at the end of the experiment, using a 5-point Likert scale (from 1= very bad to 5 = very good).

3.2. Results

3.2.1. Stage A Cooperation. We first conduct a manipulation check to verify that our Stage A game specifications worked as expected. Indeed, we observe high levels of RPD cooperation in the C-Culture treatment and low levels of RPD cooperation in the D-Culture treatment (see Figure 1). Consistent with this observation, a regression (linear probability model) predicting cooperation finds a significant positive effect of the C-Culture dummy ($p < 0.001$). A second regression shows that this difference increases as subjects learn over time (significant positive interaction between the C-Culture dummy and the decision number, $p < 0.001$, in a regression including the decision number). See Online Appendix Table A1 for regression table. Thus our Stage A manipulation successfully creates conditions of persistent cooperation (in the C-Culture) or defection (in the D-Culture).

Figure 1 (Color online) Fraction of Subjects Cooperating in Each Repeated Prisoner's Dilemma Decision of Stage A in Experiment 1, with Locally Estimated (LOESS) 95% Confidence Intervals

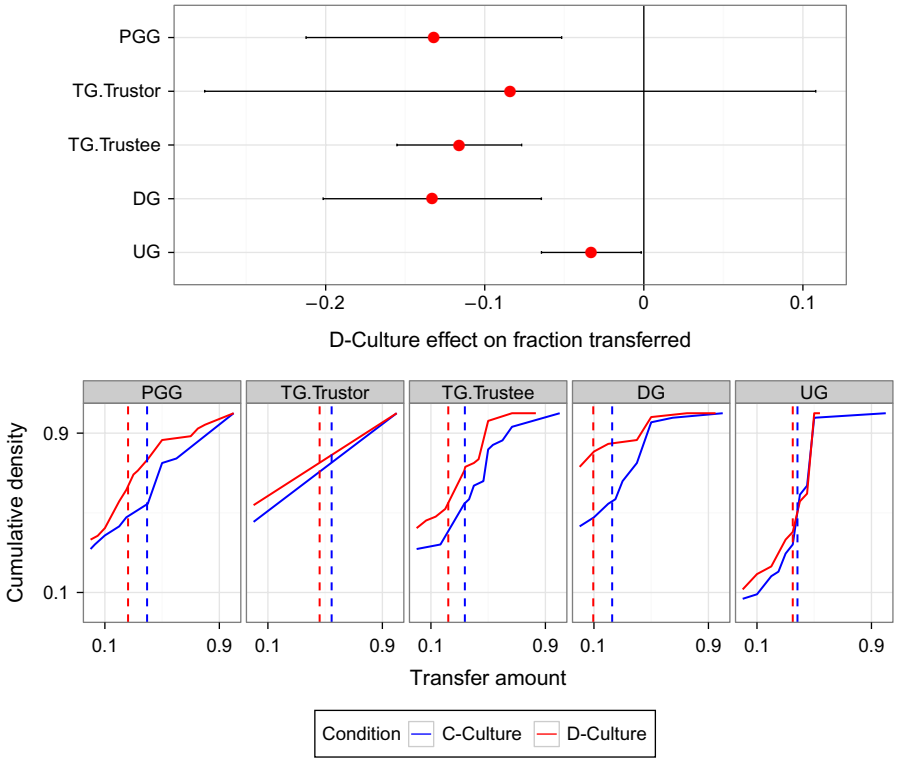


3.2.2. Stage B Prosociality. We now turn to our main question of interest: does prosociality in Stage B differ between subjects exposed to the C-Culture versus D-Culture treatments in Stage A? As predicted, we see that subjects randomized into the D-Culture treatment transfer substantially less money to others in Stage B than those randomized into the C-Culture (see Figure 2) ($p < 0.001$ with or without dummies for decision type; see Online Appendix Table A3, columns 1 and 2). When including interactions between treatment and decision type, we find only a significant interaction with the UG sender dummy ($p = 0.018$) (for all other interactions, $p > 0.4$; see Online Appendix Table A3, column 3), such that Stage A has less of an effect on UG offers. Nonetheless, Stage A still increases transfers even in the UG: when analyzing each decision separately (see Online Appendix Table A4), we find a significant positive effect in all decisions except TG trustor. (Although there is a relatively large effect size for the TG trustor, as seen in Figure 2, the standard errors are also larger than for the other decisions because TG trustors make a binary rather than scalar decision.) Thus we find strong support for our prediction that Stage A repeated-game experiences spill over to the strategically different setting of Stage B's one-shot anonymous games. Exposure to rules that cause subjects (and everyone they interact with) to cooperate or defect almost all of the time alters subsequent prosociality, with strategies that are advantageous over

[8] As the CRT has become a commonly used measure, and prior exposure to the questions may undermine their effectiveness, we use a modified version introduced in Shenhav et al. (2012), which has the following questions: (Q1) The ages of Mark and Adam add up to 28 years total. Mark is 20 years older than Adam. How many years old is Adam? (The correct answer is 4; the intuitive answer, 8.) (Q2) If it takes 10 seconds for 10 printers to print out 10 pages of paper, how many seconds will it take 50 printers to print out 50 pages of paper? (The correct answer is 10; the intuitive answer, 50.) (Q3) On a loaf of bread, there is a patch of mold. Every day, the patch doubles in size. If it takes 12 days for the patch to cover the entire loaf of bread, how many days would it take for the patch to cover half of the loaf of bread? (The correct answer is 11; the intuitive answer, 6.)

[9] Welsh et al. (2013) present evidence that performance on the CRT correlates with heuristic use to a greater extent in contexts where numerical skill is involved in arriving at the correct answer compared with contexts where numerical skill is not required. Based on these findings, CRT score should successfully tap into heuristic use in our experiment, as the games of our Stage B clearly have a numerical component. To further distinguish between general mathematical ability and reliance on intuition, we include a secondary analysis examining the number of intuitive answers provided on the CRT (rather than number of correct answers). We also note that other work has found that CRT correlations (as we use here) and actual experimental manipulations of intuitive processing have similar effects, even in nonnumerical domains (e.g., belief in God; Shenhav et al. 2012).

Figure 2 **(Color online) Mean Transfers in Stage B by Game Type**



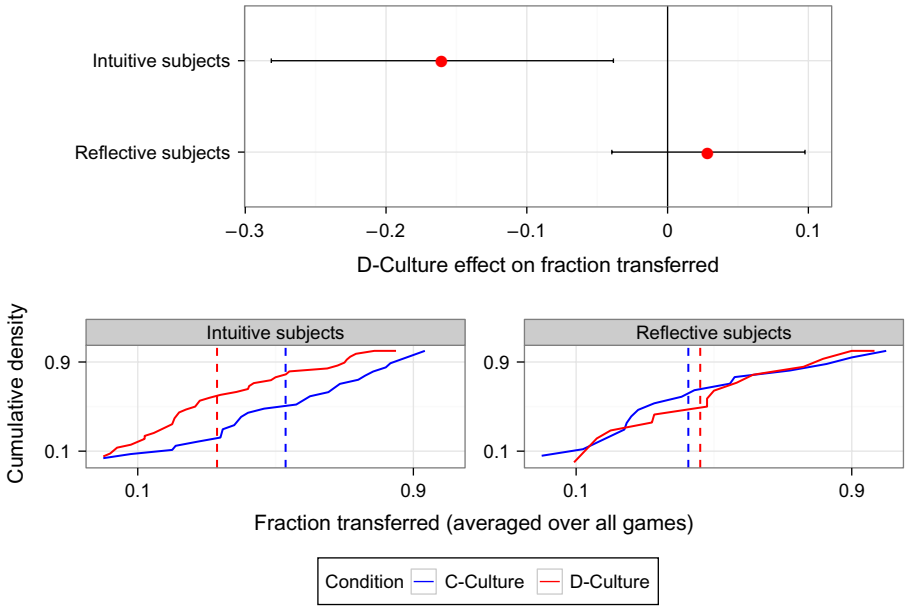
Note. Top panel: mean fraction transferred in D-Culture minus mean fraction transferred in C-Culture for each Stage B decision in Experiment 1, with error bars indicating bootstrapped 95% confidence intervals clustered on session. Bottom panel: cumulative distribution function of fraction transferred in each Stage B decision for the C-Culture (dark gray; blue online) and D-Culture (light gray; red online); mean transfers for each condition are indicated by vertical dashed lines.

many repeated trials in Stage A being carried over into the novel one-shot decision settings of Stage B.

3.2.3. Role of Heuristics in Spillovers from Stage A to Stage B. The SHH predicts that the Stage A manipulation will act largely by altering subjects' intuitive heuristic responses (rather than their rational, deliberative responses). Thus, Stage A is predicted to have less of an effect on subjects who are better at overruling their heuristic responses when making decisions (and who therefore score better on the CRT). Consistent with this prediction, we find a significant negative interaction between number of correct answers on the CRT and a C-Culture dummy when predicting Stage B prosociality ($p = 0.028$; see Online Appendix Table A5, column 1). This effect is visualized in Figure 3 by comparing the treatment effect among more intuitive thinkers (those giving one or more incorrect CRT answers: $N = 65$, $p = 0.010$; see Online Appendix Table A5, column 2) and more deliberative thinkers (those giving no incorrect CRT answers: $N = 31$, $p = 0.413$; see Online Appendix Table A5, column 3). Although the number of incorrect CRT answers is the standard measure of (non)intuitiveness, subjects might answer incorrectly by giving an answer that is neither the intuitive answer nor the correct answer (perhaps because of limited mathematical ability). Since we are particularly interested in the use of intuition (rather than general inability to do math), we note that we find a similar moderation effect using the number of specifically intuitive answers rather than just generally incorrect answers (see Online Appendix Table A6). Taken together, these results provide evidence that the changes that Stage A causes in Stage B behavior are specifically the result of reshaping intuitive heuristic responses.

3.2.4. Role of Beliefs in Spillovers from Stage A to Stage B. Can the effect of Stage A on behavior in Stage B be explained solely by changes in subjects' beliefs, rather than an actual remodeling of preferences (which are typically assumed to be fixed)? The most straightforward way that beliefs could affect behavior is by changing expectations about the decisions of one's Stage B coplayer(s). And, indeed, we find such an effect: TG trustors coming from the D-Culture expect a significantly smaller back-transfer from trustees than those from the C-Culture (33 MU versus 48 MU, $p < 0.001$; see Online Appendix Table A4, column 6). Changing expectations about the behavior of one's partner cannot, however, explain all of our results: we find treatment effects in decisions that are not influenced by such expectations.

Figure 3 **(Color online) Mean Transfers in Stage B by Subject Type**



Note. Top panel: mean fraction transferred in D-Culture minus mean fraction transferred in C-Culture averaged on over all five Stage B transfer decisions, for more intuitive subjects (those giving one or more incorrect CRT answers) and more deliberative (those giving no incorrect CRT answers) subjects; error bars indicate bootstrapped 95% confidence intervals clustered on session. Bottom panel: CDFs of average fraction transferred across all decisions for the C-Culture (dark gray; blue online) and D-Culture (light gray; red online); mean transfers for each condition are indicated by vertical dashed lines.

Trustee decisions in the TG are explicitly conditioned on their trustor's behavior (the trustee's decision is only implemented if the trustor chooses to transfer); therefore, expectations of trustor behavior have little room to influence trustee behavior. Even clearer is the DG, where the recipient makes no decision, and thus there is no room whatsoever for expectations about the partner's decision. Yet we find large and statistically significant treatment effects on both of these decisions (TG trustee: 33.7% versus 22.1% transferred, $p < 0.001$, see Online Appendix Table A4, column 4; DG: 22.7% versus 9.5% transferred, $p < 0.01$, see Online Appendix Table A4, column 1). These effects cannot be explained by beliefs about the partner's choices.

There is, however, a somewhat subtler way that beliefs might influence behavior, as captured by social preference models of type-based reciprocity (e.g., Levine 1998). These models suggest that one's desire to help another person depends on one's beliefs about how helpful that other person is. Thus, in the DG, the dictator's beliefs about the altruism of his or her recipient (e.g., what the recipient would have done had he or she been the dictator) might influence the dictator's behavior, even though the recipient makes no actual decision in the game. When making their DG decisions in Stage B, subjects do not know anything about their recipient's history of play in Stage A. They do, however, know that their recipient also participated in Stage A. Thus dictators might draw inferences about their recipient's type based on prior experience with their various Stage A partners, allowing Stage A to influence DG giving in Stage B via beliefs about type (although such inferences would likely be incorrect, since random assignment ensures that differences in Stage A play across conditions are purely a result of the rules of the game, rather than differences in the distribution of types across conditions).

Here, we present two pieces of evidence that speak against this possibility. The first comes from comparing the treatment effect on the DG versus the TG trustee. In the DG, there is ambiguity about how one's partner would have acted (leaving room for type-based beliefs to affect behavior). This is substantially less true in the TG, however, because the trustee's decision is only implemented if the trustor transfers money. Therefore, there is less ambiguity about the partner's type when choosing how much to return: trustees know that their decision will only be implemented if their partner made the "nice" choice of transferring. This implies that if the treatment effect is driven solely by changing expectations about the partner's type, we should see a substantially larger treatment effect for the DG (where there is more type ambiguity) than the TG trustee. Contrary to this prediction, however, we observe very similar effect sizes in the two games: when comparing the D-Culture and the C-Culture, 11.6% less of the stake is transferred by trustees and 13.2% less is transferred by dictators, a nonsignificant difference (interaction between treatment dummy and trustee dummy: $p = 0.68$; see

Online Appendix Table A3, column 3). Thus, reducing the amount of ambiguity regarding the partner's type does not reduce the treatment effect.

Our second piece of evidence comes from an additional supplemental experiment designed to minimize the inferences subjects draw regarding the type of their Stage B partner based on their Stage A interactions. Specifically, our supplemental experiment's Stage B consists only of a single DG where the recipient does not participate in Stage A; the DG recipients are completely passive, and the dictators are informed of this fact very prominently. Here, the play of others in Stage A should have no bearing on one's expectations regarding the Stage B recipient. We recruit 237 American subjects using the (nonstudent-based) online labor market Amazon Mechanical Turk (Amir et al. 2012, Horton et al. 2011). We then have them play an adapted version of our Stage A RPD followed by a dictator game in which the recipient has not played the RPD and receives no payment other than what is given to them in the DG. The results of this experiment closely replicate the DG results observed in Experiment 1, with nearly twice as much DG giving after the C-Culture compared with the D-Culture (27.4% versus 16.5% transferred, rank-sum $p < 0.001$). See Online Appendix §4 for details. Thus reducing the grounds for using Stage A to make inferences about the type of the DG recipient in Stage B does not appreciably undermine the treatment effect.

In sum, although we cannot completely rule out the possibility that our treatment effect is driven by beliefs about the distribution of types, we provide evidence that is inconsistent with this possibility and that instead points to an actual remodeling of preferences (e.g., internalized notions of appropriate behavior).

3.2.5. Role of Mood in Spillovers from Stage A to Stage B. Finally, we note that our treatment effect does not appear to be driven by mood. Most simply, we find little difference in our 1-to-5 mood measure between the C-Culture (mean = 3.45, sd = 0.588) and the D-Culture (mean = 3.29, sd = 0.690; rank-sum, $p = 0.38$) in the three sessions of Experiment 1 in which mood was measured after Stage B. (We also find no significant effect of condition on mood in the nine sessions of Experiment 2 where mood was measured—C-Culture: mean = 3.43, sd = 0.095; D-Culture: mean = 3.36, sd = 0.106; rank-sum, $p = 0.46$). Further evidence comes from replicating our main analysis using only these three sessions (see Online Appendix Table A7). If the effect of treatment is driven by Stage A altering subjects' mood, then the treatment coefficient will be smaller when controlling for mood. On the contrary, however, we find that the coefficient on treatment *increases* in magnitude when including mood in the regressions (C-Culture dummy: coefficient = 0.128 without mood and 0.139 with mood). Thus we do not find evidence that our treatment effect is driven by mood. We do note, however, that mood was measured after Stage B decisions were made, and thus it is not a pure measure of Stage A's effect on mood. Further investigation of the role of mood in our treatment effect is a worthwhile direction for future work.

4. Experiment 2: Effects of Cooperative Environment on Norm Enforcement

4.1. Experimental Design

A key element of norms is not just the desire to act in a particular way but also the willingness to sanction those who do not act accordingly. In Experiment 2, we therefore ask whether Stage A also effects subjects' punishment of selfishness (i.e., willingness to pay to reduce the payoffs of noncooperators). To do so, we follow Stage A with a battery of one-shot punishment games, played using the strategy method with no feedback: a DG with third-party punishment (3PDG), a prisoner's dilemma with third-party punishment and reward (3PPD), and a PGG with punishment (PGP). In each of these punishment decisions, players face a different type of dilemma from Experiment 1: they have the chance to decrease the payoff of one or more others at a cost to themselves, based on the others' behavior.

In the 3PDG game, subjects are matched in groups of three. One subject is assigned to be the dictator and unilaterally decides on a split of 100 MU between herself and a second subject assigned to be the recipient. A third subject, the sanctioner, is given an endowment of 100 MU and indicates how many MU (up to 20) she would spend to reduce the dictator's payoff, depending on the dictator's chosen division.[10] Each monetary unit spent by the third party reduces the dictator's payoff by 5 MU.

In the 3PPD game, subjects are matched into groups of four. Two subjects are selected to be PD players and play a one-shot binary-choice PD with each other using the following payoff matrix (in MU):

	C	D
C	80, 80	0, 120
D	120, 0	20, 20

The other two subjects are sanctioners and are each given an endowment of 100 MU. The sanctioners choose how much (up to a maximum of 20 MU) to

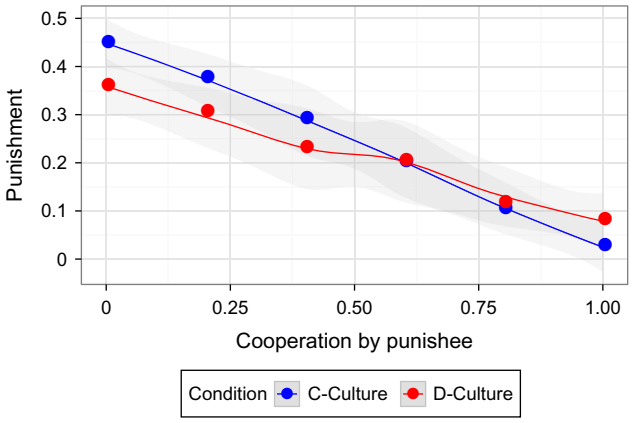
[10] The dictator had the option of transferring 0, 10, 20, 30, 40, or 50 MU. The sanctioner indicated the number of MU to spend on reducing the dictator's payoff for each of these possible transfer options.

spend on reducing one of the two PD players' payoff, based on that player's decision in the PD. We also use this game to address concerns that have been raised about third-party punishment experiments where punishment was the only option (Pedersen et al. 2013): we allow third parties to reward as well as punish the PD players. Each monetary unit spent by the third party on punishing reduces the PD player's payoff by 5 MU, and each monetary unit spent on rewarding increases the PD player's payoff by 5 MU.

In the PGP game, subjects are matched into groups of four and play the same PGG as in Stage B of Experiment 1 (100 MU endowment, efficiency factor of 1.6). Unlike Experiment 1, however, after making a PGG decision, each subject has the opportunity to pay up to 20 MU to sanction other group members based on contribution amount.[11] Each monetary unit spent reduces the sanctioned player's payoff by 5 MU.

In all three games, in addition to punishment decisions, subjects are asked to rate how "socially inappropriate" each possible punishee behavior was, using a 7-point Likert scale.

Figure 4 **(Color online) Third-Party Punishment in the 3PDG and 3PPD Games of Stage B in Experiment 2, as a Function of the Punishee's Cooperativeness (Fraction of Punishment Endowment Spent, Averaged Across the 3PDG and 3PPD Games), with LOESS 95% Confidence Intervals**



Note. Punishee's cooperativeness for the 3PDG is normalized such that giving nothing corresponds to cooperativeness 0 and giving 50% corresponds to cooperativeness 1, and for the 3PPD, such that defecting corresponds to cooperativeness 0 and cooperating to cooperativeness 1.

4.2. Results

4.2.1. Stage A Cooperation. As in Experiment 1, we effectively induce substantially more Stage A cooperation in the C-Culture treatment compared with the D-Culture treatment (58% C versus 22% C, $p < 0.001$; see Online Appendix Table A11).

4.2.2. Stage B Third-Party Punishment. We begin by considering the effect of Stage A on third-party punishment in Stage B (see Figure 4). In the third-party punishment games (3PDG and 3PPD), punishment is impartial (as the punisher is not affected by the behavior of the punishee) and thus reflects pure norm enforcement (Fehr and Fischbacher 2004b).

To test whether Stage A affected norm enforcement in Stage B, we follow the approach of Experiment 1 and analyze the data from the 3PPD and the 3PDG together in a single regression (predicting punishment as a function of the punishee's level of cooperation). To make cooperation in the DG and PD comparable, we normalize such that maximum selfishness (transferring nothing in DG, playing D in PD) has a value of 0 and maximum prosociality (transferring 50 in the DG, playing C in PD) has a value of 1. Regressing the amount of punishment on the punishee's level of cooperation, a dummy for C-Culture, and an interaction between punishee's level of cooperation and the C-Culture dummy (Online Appendix Table A12, column 1) shows a significant negative main effect of punishee's cooperation (more cooperative actions were punished less, $p < 0.001$), a significant positive main effect of C-Culture (a maximally selfish action was punished more in the C-Culture than in the D-Culture, $p = 0.021$) and a significant negative interaction (for a given increase in punishee's cooperation, punishment decreased more in the C-Culture than in the D-Culture; that is, punishment was more selectively targeted at selfish behavior in the C-Culture, $p = 0.006$).

These results are robust to including fixed effects for game type (see Online Appendix Table A12, column 2) as well as interactions between game type and treatment (see Online Appendix Table A12, column 3) or to analyzing the two games separately (see Online Appendix Figure A3 and Table A12, columns 4 and 5). We find that subjects' ratings of the inappropriateness of DG sending and PD cooperation decisions are affected by the treatment in a qualitatively similar way (i.e., positive C-Culture dummy coefficient, negative interaction between C-Culture and punishee prosociality) but that these self-report measures are less sensitive than actual punishment, and statistical significance is not achieved (see Online Appendix Table A13). We also find that the treatment does not have a significant effect on rewarding in the 3PPD (see Online Appendix Table A12, column 6).

[11] In the PGP contribution phase, subjects can choose one of the following contribution amounts: 0, 25, 50, 75, or 100. In the PGP punishment phase, subjects are paired up and indicate how many MU to spend punishing the person with whom they are paired, for each possible partner contribution amount. We do not offer a reward option in the PGP (as in Rand et al. 2009), but based on the similarity of results between of our third-party punishment games with and without reward, we think it is unlikely that adding a reward option to the PGP would change our results substantially.

Importantly, we find significant treatment effects on third-party punishment among more intuitive subjects but not among more deliberative subjects (see Online Appendix Table A14), replicating the pattern seen for prosociality in Experiment 1.

These results demonstrate that Stage A altered subjects' conception of cooperation as a norm to be enforced as well as followed, rather than just priming subjects to themselves give or not: subjects in the C-Culture condition not only behaved more cooperatively in Experiment 1 but were also more willing to pay to punish noncooperators in Experiment 2.

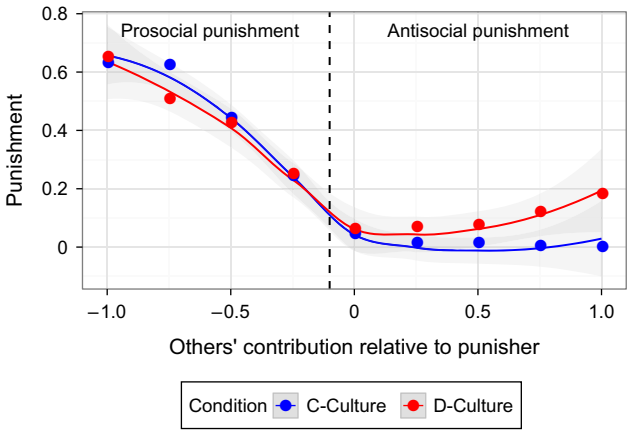
4.2.3. Stage B Public Goods Game with Punishment. We now turn to the effect of Stage A on punishment in the public goods game. Unlike the 3PDG and 3PPD games, punishment in the public goods game is not wholly impartial: noncontributors have a negative impact on the earnings of the potential punisher as well as on other group members. Thus motives other than norm enforcement, such as retaliation or the desire to outearn other group members, may also drive punishment in this setting (Bolton and Ockenfels 2000, Ellingsen et al. 2012, Espín et al. 2012, Herrmann et al. 2008, Rand and Nowak 2011).

Perhaps as a result of these mixed motives, cross-cultural experiments using the public goods game (e.g., Herrmann et al. 2008) have found relatively little variation in prosocial punishment (punishment of those contributing less than the punisher, or PSP). The level of "antisocial punishment" (punishment of those contributing as much or more than the punisher, or ASP), conversely, has been shown to vary dramatically across cultures. Thus cross-cultural effects on PGG punishment vary based on the difference between the punisher's contribution and the punishee's contribution. We therefore consider Stage A effects on PGG punishment based on this difference in punisher and punishee contribution levels (see Figure 5).

First, we consider PSP of those who contribute less than the punisher (the left half of Figure 5). Consistent with previous cross-cultural results (Herrmann et al. 2008), we find no significant difference in overall PSP between our C-Culture and D-Culture treatments (see Online Appendix Table A15, column 1, $p = 0.72$) and no interaction between treatment and the difference in punisher's and punishee's contribution amount (see Online Appendix Table A15, column 2, $p = 0.65$).

Next, we turn to ASP targeted at those who contributed as much or more than the punisher (the right half of Figure 5). Our results are again consistent with the cross-cultural work of (Herrmann et al. 2008), where such punishment is rare among Western college students but occurs frequently in countries with weaker institutions: we find virtually no antisocial punishment in the C-Culture treatment, but some begins to appear in the D-Culture treatment.

Figure 5 **(Color online) Punishment in the PGP of Stage B in Experiment 2, as a Function of the Difference Between the Punishee's Normalized PGG Contribution and the Punisher's Normalized PGG Contribution (with LOESS 95% Confidence Intervals)**



Note. Negative x-axis values correspond to prosocial punishment (where the punisher contributes more than the punishee), whereas zero or positive x-axis values correspond to antisocial punishment (where the punishee contributes as much or more than the punisher).

There is a trend in the direction of less overall ASP in the C-Culture (see Online Appendix Table A15, column 3, $p = 0.16$), and there is a marginally significant interaction between the C-Culture dummy and contribution difference (see Online Appendix Table A15, column 4, $p = 0.098$) such that in the D-Culture, subjects are somewhat more likely to punish as the punishee's contribution *increases*.

These results provide further evidence that our Stage A manipulation is altering the norms applied in Stage B, and they hint that cross-cultural differences in ASP (Ellingsen et al. 2012, Herrmann et al. 2008) may in part be explained by norms developed in an environment where cooperative behavior is not an advantageous strategy (a possibility that deserves further exploration in future research).

5. Beyond Play in Economic Games

Finally, we ask whether the effects of Stage A extend beyond play in economic games. To do so, we employ the World Values Survey (WVS), one of the most widely used instruments for the measure of differences in cultural norms (Gächter et al. 2004, Putnam 2000). A version of the World Values Survey question measuring generalized trust was included in the postexperimental survey in each of our experiments:[12]

[12] Certain subjects' questionnaires did not include this question, but not in a systemic way: three sessions of Experiment 1 contained the trust question (48 observations), all but one session from the Experiment 2 contained the trust question (109 observations), and all 237 observations from the supplemental experiment described in the online appendix included the trust question.

subjects were asked, "How much do you agree with the statement: 'Most people can be trusted'?" and indicated their response using a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). This specific continuous implementation is used by Peysakhovich et al. (2014), who find that agreement with the statement correlates strongly with prosociality in the one-shot anonymous PGG and DG. The correlation with DG giving suggests that this question measures more than simple beliefs about the behavior of others that is implied by the word "trust" but instead is a more general measure of the responder's prosociality. Additional evidence for this broader interpretation of the WVS trust question comes from Glaeser et al. (2000), who find that it predicts trustworthiness much better than trust. Thus, it seems likely that the Stage A effects on prosocial norms observed in the game behavior of Experiments 1 and 2 may also translate into effects on this generalized trust measure.

Indeed, subjects randomized into the D-Culture report significantly lower levels of generalized trust than those randomized into the C-Culture, both overall (see Online Appendix Table A16, column 1, $p < 0.001$) and among the lab (Online Appendix Table A16, column 3, $p = 0.040$) and online subjects (Online Appendix Table A16, column 4, $p = 0.001$) separately. Furthermore, we find evidence of a similar CRT moderation effect to that observed with the games: Stage A significantly changes WVS trust among more intuitive subjects, but not among more deliberative subjects (see Online Appendix Table A17).

Thus, the effect of our Stage A manipulation is not restricted to economic games; we also recreate variation in a standard survey instrument used to measure culture through random assignment to an interaction environment in the lab. With this measure, the heuristics developed in Stage A are misapplied not to behavior in one-shot anonymous games, but rather to a survey question that combines assessments about how trustworthy others are with one's own level of prosociality.

6. Concluding Discussion

Here, we have shown that externally imposed interaction rules can dramatically alter subjects' internalized norms: immersing subjects in an environment that incentivizes cooperation or defection for less than 20 minutes leads to large differences in subsequent one-shot anonymous prosociality and sanctioning, as well as generalized trust. Furthermore, these effects are driven by subjects that rely on heuristics and virtually disappear among highly deliberative subjects. Our findings demonstrate the power of spillover effects for shaping behavior in one-shot anonymous settings. Furthermore, we shed light on the key role played by intuition and heuristics in this internalization of cooperative norms (as proposed by the social heuristics hypothesis; see Rand et al. 2014b). In doing so, we help to explain why people often cooperate in one-shot anonymous contexts and why such behavior might vary across organizations and cultures. More broadly, our results demonstrate that laboratory experiments can be a powerful tool for studying culture and the interplay between social environments and internalized norms.

Our results cannot be explained by changes in beliefs regarding the decisions of coplayers, as we see from the treatment effects in the dictator game and the trust game trustee. We also present evidence that our treatment effects are not easily explained by changes in beliefs regarding the coplayer's type (see §3.2.4), although we cannot definitively rule out this possibility. We argue that our results therefore suggest that social *preferences* may be remodeled on relatively short time scales (20 minutes or less of RPD play).[13] This suggestion runs counter to the standard assumption in economic models that preferences are fixed, and it points to the importance of considering both where social preferences come from and how they change. These questions are not addressed by most social preferences models, which take preferences as given (Bolton and Ockenfels 2000, Charness and Rabin 2002, Fehr and Schmidt 1999, Levine 1998, Rabin 1993). In particular, the evidence we present for the role of heuristics in this remodeling of preferences suggests that adding a dual process perspective to social preference models (e.g., Dreber et al. 2014, Loewenstein and O'Donoghue 2004) is important when exploring preference change. Evolutionary game-theoretic models *do* often try to shed light on the issues of origin and change of prosocial behavior (e.g., Alger and Weibull 2010, Bowles 2001, Dekel et al. 2007, Rand and Nowak 2012, Rand et al. 2013, Samuelson 2001) but often without explicitly connecting to specific models of social preferences. Extending social preference models to incorporate heuristic processing and preference change and more directly linking evolutionary models to models of social preferences are important directions for future theoretical work.

It is also interesting to consider how our results relate to the economic theory literature on habit formation, growing out of the seminal work of Becker

[13] Convergent empirical evidence for such remodeling comes from studies showing that short-run factors such as recent experiences or organizational memberships can affect internalized behaviors related to trust and cooperation (e.g., Alesina and La Ferrara 2002, Bellows and Miguel 2009, Fisman and Khanna 1999).

and Murphy (1988). The Becker–Murphy model defines habit formation as adaptation to levels of consumption of a particular good: the amount of the good one desires to consume in the current period is increasing in the amount one consumed in the previous period. To apply this model to our experiments, think of prosociality as the good being consumed. To explain the spillover over effects we observe in Experiment 1, we would have to assume that subjects think of RPD cooperation in Stage A and money transfers in the one-shot games of Stage B as instances of the same "prosociality" good: subjects in the C-Culture consume more prosociality in the Stage A RPD than those in the D-Culture, and they therefore are more inclined to consume prosociality in the anonymous one-shot games of Stage B. This assumption may be reasonable, given evidence that an individual's play correlates strongly across different prosociality games, suggesting that a common preference drives behavior across these games (Peysakhovich et al. 2014). To explain Experiment 2, we would have to further assume that subjects experience Stage A RPD cooperation and Stage B third-party punishment of selfishness as being the same good. This assumption is contradicted, however, by evidence that an individual's play in cooperation games is *not* predictive of his or her third-party punishment behavior (Peysakhovich et al. 2014). This suggests that cooperation and punishment are psychologically distinct goods and that our results are driven by norm internalization rather than by pure habituation to cooperation.

Although we demonstrate change in behavior following relatively brief exposure to cooperative or noncooperative environments, we do not systematically vary treatment durations. Thus we cannot estimate how different lengths of exposure to the treatment translate into different levels of change in subsequent behavior. Creating such a "dose-response" curve is an important direction for future experimental work. So too is exploring the extent to which the Stage A treatment generalizes across contexts. We provide some evidence of generalization, in that RPD behavior generalizes to other prosociality games, to punishment games, and to the non-game survey measure of trust from the World Values Survey. We do not contend, however, that our Stage A treatment erases a lifetime of previous experience and cultural context, or irrevocably changes our subjects' preferences. Rather, our experiments should be seen as a proof-of-concept that the behaviors cultivated by particular environments travel beyond those situations and that cooperation includes relatively malleable components. Future work should evaluate how far beyond the experimental context our effect extends, as well as how our artificial Stage A can be translated into more contextualized settings (which may lead to even broader generalization).

A related question concerns the persistence of the effect we observe. Some insight comes from the findings of Duffy and Ochs (2009) and Fréchette and Yuksel (2013), who examine play in a series of RPD games: in a first phase, subjects played RPDs where cooperation is an equilibrium; then in a second stage, the game specification is modified such that cooperation is not an equilibrium.[14] Both studies find a large reduction in cooperation when switching from the first stage to the second. Consistent with our findings, however, there is some indication of spillover: cooperation in the first decision of the second stage is higher than in later periods, and a process of decay occurs before aggregate behavior stabilizes at an extremely low level. This rapid decay indicates that subjects quickly adapt to their new environment. However, this does not mean that these spillover effects are trivial, for two reasons. First, in real-world applications, environmental conditions are themselves typically highly persistent: people interact every day in an environment that favors either cooperation or defection. Thus they "receive treatment" constantly, supporting a continual spillover of norms into the subset of interactions that are one-shot and anonymous. Second, many situations of economic interest have multiple equilibria (unlike our Stage B games or those of Duffy and Ochs 2009 and Fréchette and Yuksel 2013). Thus even short-lived spillover effects can have long-lasting impacts by changing the initial conditions and shifting between basins of attraction of different equilibria, as demonstrated in the context of weak-link coordination games by Brandts and Cooper (2006). Quantifying the persistence of our Stage A effect and its consequences in settings with multiple equilibria is an important direction for future experiments.

In our experiments, we modeled environments that make cooperation advantageous or disadvantageous using the framework of repeated games. We used repetition because it is a well-established paradigm that creates future consequences for today's actions. Critically, however, we are not claiming that repetition *itself* varies across cultures. Instead, we are arguing that variation in whether institutional incentives support cooperative equilibria (*modeled* in our experiments by repetition) leads to variation in internalized norms. Other work on path-dependent preferences suggests that the specific form taken by these institutional incentives also matters. For example, Bohnet and Baytelman (2007) find that adding communication, punishment, or fixed repetition to the TG

[14] This switch is accomplished by switching from fixed matching to random matching in the case of Duffy and Ochs (2009) and by changing the stage game payoffs in the case of Fréchette and Yuksel (2013).

increases trust and trustworthiness almost exclusively as a result of changes in expectations (and if anything, these cooperation-inducing institutions "crowd out" intrinsic preference-based trust, rather than positively influencing social preferences as in our experiment). Bohnet and Huck (2004) find no differences in one-shot TG play based on whether subjects previously played a series of stranger-matched (i.e., one-shot) TGs or repeated play, and they find some evidence that stranger-matched play with a reputation system actually decreases subsequent trust; in a sequential step-level PGG, Cooper and Stockman (2011) find relatively little effect of prior experience under rules that emphasized different kinds of equity concerns, and whatever effects they did find were very short-lived. And Herz and Taubinsky (2013) find that prior experience in markets with substantial competitive pressure results in a large and persistent decrease in fairness concerns in the UG. Given this heterogeneity, future work directly comparing the effect of different environmental structures that incentivize cooperative behavior (e.g., institutions such as markets, democratic governance, or centralized reward and punishment) is needed.

We also do not claim that the arrow from institutionally created environments to preferences goes only in one direction—the coevolution of norms and interaction environments (e.g., institutions) involves a feedback loop between the two. In situations where institutions (e.g., Stage A rules) work by creating cooperative equilibria, noncooperative equilibria often also exist.[15] Thus the effectiveness of an institution is in part determined by the norms of those whom the institution governs. Multiple equilibria can cause norm persistence (rather than the malleability seen in our studies), as demonstrated, for example, by Nunn and Wantchekon (2011), who show that individuals whose ancestors come from areas in Africa that had been more affected by the slave trade continue to have weaker levels of interpersonal trust today.[16] Future laboratory studies should examine how institutional rules and baseline norms interact. An important part of such work will involve cross-cultural studies, where our Stage A manipulation is applied to groups with differing baseline levels of cooperation. Such studies will shed light on whether the treatment effect we observe is driven by the C-Culture increasing cooperation, the D-Culture decreasing cooperation, or both. Based on the SHH, we predict that baseline behavior in one-shot anonymous games will resemble Stage B of the C-Culture treatment in places with strong institutions, such as the United States, and will resemble Stage B of the D-Culture treatment in places with weak institutions (e.g., see Gächter et al. 2010).

Research in the social and behavioral sciences is increasingly focusing not just on generating insights into the basic science underlying human behavior but also on applying these insights to institutions, markets, incentives, and organizations outside the laboratory (Fudenberg and Peysakhovich 2014, Gerber and Rogers 2009, Gneezy and List 2006, Kraft-Todd et al. 2015, Rand et al. 2014a, Roth 2002, Thaler and Benartzi 2004, Thaler and Sunstein 2008, Yoeli et al. 2013). The experiments presented here have clear practical implications for building cooperative cultures in organizations. They suggest that cross-organizational differences in culture are in large part determined by cross-organizational differences in which behaviors (cooperative or noncooperative) are rewarded. These differences in optimal behavior are strongly influenced by the organization's institutions (i.e., how incentives are structured, the focus of the current paper). However, the way in which people are selected to join the organization also has an important effect on organizational culture: as discussed above, the right incentives can create cooperative equilibria, but groups can still coordinate on noncooperative equilibria if many of the people entering the group are initially noncooperative. Thus it is ideal for organizations to both design effective institutions and (at least to some extent) avoid hiring noncooperative individuals.

Taken together, the experiments we present here demonstrate the power of previous experience for shaping our behavior in one-shot anonymous settings. They also open the door for a wide array of possible applications for organizations interested in increasing the incidence of cooperative behavior. Prosociality induced by environmental constraints can have a dramatic influence on behavior even in situations where those constraints do not apply.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/mnsc.2015.2168.

Acknowledgments

The authors thank Robert Aumann, Colin Camerer, John Clithero, Armin Falk, Guillaume Frechette, Drew Fudenberg, Ed Glaeser, Joseph Henrich, Benedikt Herrmann, Moshe Hoffman, Jillian Jordan, David Laibson, Martin Nowak, Nathan Nunn, Pietro Ortoleva, Aurelie Ouss, Antonio Rangel, Peter Richerson, Daria Roithmayer, Al Roth, Klaus Schmidt, Dmitry Taubinsky, Julian Wills; members of

[15] Note that here we discuss multiple equilibria in the context of Stage A, rather than above where we discuss possible effects of multiple equilibria in the Stage B game.

[16] Existing theories for such persistence point to channels such as parents "instilling" values in their children (e.g., Bisin and Verdier 2000, Tabellini 2008).

the Rangel Lab; and seminar participants at Harvard, Princeton, Yale, Massachusetts Institute of Technology, Brown, and Stony Brook for their invaluable comments. They also thank the John Templeton Foundation for providing funding for this work. The authors declare no conflict of interest.

References

Alesina A, La Ferrara E (2002) Who trusts others? *J. Public Econom.* 85(2):207–234.

Alger I, Weibull JW (2010) Kinship, incentives, and evolution. *Amer. Econom. Rev.* 100(4):1725–1758.

Amir O, Rand DG, Gal YK (2012) Economic games on the internet: The effect of $1 stakes. *PLoS ONE* 7(2):e31461.

Becker GS, Murphy KM (1988) A theory of rational addiction. *J. Political Econom.* 96(4):675–700.

Bellows J, Miguel E (2009) War and local collective action in Sierra Leone. *J. Public Econom.* 93(11):1144–1157.

Bisin A, Verdier T (2000) "Beyond the melting pot": Cultural transmission, marriage, and the evolution of ethnic and religious traits. *Quart. J. Econom.* 115(3):955–988.

Blonski M, Ockenfels P, Spagnolo G (2011) Equilibrium selection in the repeated prisoner's dilemma: Axiomatic approach and experimental evidence. *Amer. Econom. J.: Microeconom.* 3(3): 164–192.

Bohnet I, Baytelman Y (2007) Institutions and trust implications for preferences, beliefs and behavior. *Rationality Soc.* 19(1):99–135.

Bohnet I, Huck S (2004) Repetition and reputation: Implications for trust and trustworthiness when institutions change. *Amer. Econom. Rev.* 94(2):362–366.

Bolton GE, Ockenfels A (2000) ERC: A theory of equity, reciprocity, and competition. *Amer. Econom. Rev.* 90(1):166–193.

Bowles S (1998) Endogenous preferences: The cultural consequences of markets and other economic institutions. *J. Econom. Literature* 36(1):75–111.

Bowles S (2001) Individual interactions, group conflicts, and the evolution of preferences. Durlauf SN, Young HP, eds. *Social Dynamics* (MIT Press, Cambridge, MA), 155–190.

Bowles S, Gintis H (2002) Prosocial emotions. Blume LE, Durlauf SN, eds. *The Economy as an Evolving Complex System, III: Current Perspectives and Future Directions* (Oxford University Press, Oxford, UK), 339–363.

Bowles S, Gintis H (2003) Origins of human cooperation. Hammerstein P, ed. *Genetic and Cultural Evolution of Cooperation* (MIT Press, Cambridge, MA), 429–443.

Boyd R, Richerson PJ (2009) Culture and the evolution of human cooperation. *Philosophical Trans. Roy. Soc. B: Biol. Sci.* 364(1533): 3281–3288.

Brandts J, Cooper DJ (2006) A change would do you good.... An experimental study on how to overcome coordination failure in organizations. *Amer. Econom. Rev.* 96(3):669–693.

Camerer CF, Fehr E (2002) Measuring social norms and preferences using experimental games: A guide for social scientists. Working paper, California Institute of Technology, Pasadena.

Cappelen AW, Moene KO, Sørensen EØ, Tungodden B (2013) Needs versus entitlements—An international fairness experiment. *J. Eur. Econom. Assoc.* 11(3):574–598.

Charness G, Rabin M (2002) Understanding social preferences with simple tests. *Quart. J. Econom.* 117(3):817–869.

Chudek M, Henrich J (2011) Culture gene coevolution, norm-psychology and the emergence of human prosociality. *Trends Cognitive Sci.* 15(5):218–226.

Cone J, Rand DG (2014) Time pressure increases cooperation in competitively framed social dilemmas. *PLoS ONE* 9(12): e115756.

Cooper DJ, Stockman CK (2011) History dependence and the formation of social preferences: An experimental study. *Econom. Inquiry* 49(2):540–563.

Cornelissen G, Dewitte S, Warlop L (2011) Are social value orientations expressed automatically? Decision making in the dictator game. *Personality Soc. Psych. Bull.* 37(8):1080–1090.

Dal Bó P (2005) Cooperation under the shadow of the future: Experimental evidence from infinitely repeated games. *Amer. Econom. Rev.* 95(5):1591–1604.

Dal Bó P, Fréchette GR (2011) The evolution of cooperation in infinitely repeated games: Experimental evidence. *Amer. Econom. Rev.* 101(1):411–429.

Dekel E, Ely JC, Yilankaya O (2007) Evolution of preferences. *Rev. Econom. Stud.* 74(3):685–704.

Dreber A, Fudenberg D, Levine DK, Rand DG (2014) Self-control, social preferences and the effect of delayed payments. Working paper, Washington Unversity in St. Louis, St. Louis. http://ssrn.com/abstract=2477454.

Dreber A, Rand DG, Fudenberg D, Nowak MA (2008) Winners don't punish. *Nature* 452(7185):348–351.

Duffy J, Ochs J (2009) Cooperative behavior and the frequency of social interaction. *Games Econom. Behav.* 66(2):785–812.

Ellingsen T, Herrmann B, Nowak MA, Rand DG, Tarnita CE (2012) Civic capital in two cultures: The nature of cooperation in Romania and USA. Working paper, Stockholm School of Economics, Stockholm. http://ssrn.com/abstract=2179575.

Epstein S, Pacini R, Denes-Raj V, Heier H (1996) Individual differences in intuitive–experiential and analytical–rational thinking styles. *J. Personality Soc. Psych.* 71(2):390–405.

Espín AM, Brañas-Garza P, Herrmann B, Gamella JF (2012) Patient and impatient punishers of free-riders. *Proc. Roy. Soc. Ser. B* 279(1749):4923–4928.

Fehr E, Fischbacher U (2004a) Social norms and human cooperation. *Trends Cognitive Sci.* 8(4):185–190.

Fehr E, Fischbacher U (2004b) Third-party punishment and social norms. *Evolution Human Behav.* 25(2):63–87.

Fehr E, Gächter S (2000) Cooperation and punishment in public goods experiments. *Amer. Econom. Rev.* 90(4):980–994.

Fehr E, Schmidt K (1999) A theory of fairness, competition and cooperation. *Quart. J. Econom.* 114(3):817–868.

Fischbacher U (2007) z-Tree: Zurich toolbox for ready-made economic experiments. *Experiment. Econom.* 10(2):171–178.

Fisman R, Khanna T (1999) Is trust a historical residue? Information flows and trust levels. *J. Econom. Behav. Organ.* 38(1):79–92.

Fréchette GR, Yuksel S (2013) Infinitely repeated games in the laboratory: Four perspectives on discounting and random termination. Working paper, New York University, New York. http://ssrn.com/abstract=2225331.

Frederick S (2005) Cognitive reflection and decision making. *J. Econom. Perspect.* 19(4):25–42.

Fudenberg D, Maskin ES (1986) The folk theorem in repeated games with discounting or with incomplete information. *Econometrica* 54(3):533–554.

Fudenberg D, Maskin ES (1990) Evolution and cooperation in noisy repeated games. *Amer. Econom. Rev.* 80(2):274–279.

Fudenberg D, Peysakhovich A (2014) Recency, records and recaps: Learning and non-equilibrium behavior in a simple decision problem. *Proc. 15th ACM Conf. Econom. Computation* (ACM, New York), 971–986.

Fudenberg D, Rand DG, Dreber A (2012) Slow to anger and fast to forgive: Cooperation in an uncertain world. *Amer. Econom. Rev.* 102(2):720–749.

Gächter S, Herrmann B, Thöni C (2004) Trust, voluntary cooperation, and socio-economic background: Survey and experimental evidence. *J. Econom. Behav. Organ.* 55(4):505–531.

Gächter S, Herrmann B, Thöni C (2010) Culture and cooperation. *Philosophical Trans. Roy. Soc. B: Biol. Sci.* 365(1553):2651–2661.

Gerber AS, Rogers T (2009) Descriptive social norms and motivation to vote: Everybody's voting and so should you. *J. Politics* 71(1):178–191.

Gigerenzer G, Goldstein DG (1996) Reasoning the fast and frugal way: Models of bounded rationality. *Psych. Rev.* 103(4):650.

Gigerenzer G, Todd PM, Group AR (1999) *Simple Heuristics That Make Us Smart* (Oxford University Press, Oxford, UK).

Gilovich T, Griffin D, Kahneman D (2002) *Heuristics and Biases: The Psychology of Intuitive Judgment* (Cambridge University Press, Cambridge, UK).

Gintis H (2003) The hitchhiker's guide to altruism: Gene-culture coevolution, and the internalization of norms. *J. Theoret. Biol.* 220(4):407–418.

Glaeser EL, Laibson DI, Scheinkman JA, Soutter CL (2000) Measuring trust. *Quart. J. Econom.* 115(3):811–846.

Gneezy U, List JA (2006) Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica* 74(5):1365–1384.

Hauge KE, Brekke KA, Johansson L-O, Johansson-Stenman O, Svedsäter H (2015) Keeping others in our mind or in our heart? Distribution games under cognitive load. *Experimental Econom.*, ePub ahead of print June 25, http://dx.doi.org/10.1007/s10683-015-9454-z.

Henrich J, Boyd R, Bowles S, Camerer C, Fehr E, Gintis H, McElreath R, et al. (2005) "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral Brain Sci.* 28(6):795–855.

Henrich J, McElreath R, Barr A, Ensminger J, Barrett C, Bolyanatz A, Cardenas JC, et al. (2006) Costly punishment across human societies. *Science* 312(5781):1767–1770.

Henrich J, Ensminger J, McElreath R, Barr A, Barrett C, Bolyanatz A, Cardenas JC, et al. (2010) Markets, religion, community size, and the evolution of fairness and punishment. *Science* 327(5972):1480–1484.

Herrmann B, Thoni C, Gächter S (2008) Antisocial punishment across societies. *Science* 319(5868):1362–1367.

Herz H, Taubinsky D (2013) Market experience is a reference point in judgments of fairness. Working paper, Harvard University, Cambridge, MA. http://ssrn.com/abstract=2297773.

Horton JJ, Rand DG, Zeckhauser RJ (2011) The online laboratory: Conducting experiments in a real labor market. *Experiment. Econom.* 14(3):399–425.

Jordan JJ, McAuliffe K, Rand DG (2014) The effects of endowment size and strategy method on third-party punishment. Working paper, Yale University, New Haven, CT. http://ssrn.com/abstract=2427274.

Jordan JJ, Peysakhovich A, Rand DG (2015) Why we cooperate. Decety J, Wheatley T, eds. *The Moral Brain: A Multidisciplinary Perspective* (MIT Press, Cambridge, MA), 87–104.

Kahneman D (2003) A perspective on judgment and choice: Mapping bounded rationality. *Amer. Psychologist* 58(9):697–720.

Kraft-Todd G, Yoeli E, Bhanot S, Rand D (2015) Promoting cooperation in the field. *Current Opinion in Behavioral Sci.* 3:96–101.

La Porta R, Lopez-de-Silanes F, Shleifer A, Vishny RW (2001) Trust in large organizations. Dasgupta P, Serageldin I, eds. *Social Capital: A Multifaceted Perspective* (World Bank Publications, Washington, DC), 310–321.

Leana CR III, Van Buren HJ (1999) Organizational social capital and employment practices. *Acad. Management Rev.* 24(3):538–555.

Levine DK (1998) Modeling altruism and spitefulness in experiments. *Rev. Econom. Dynam.* 1(3):593–622.

Loewenstein GF, O'Donoghue T (2004) Animal spirits: Affective and deliberative processes in economic behavior. Working paper, Carnegie Mellon University, Pittsburgh. http://ssrn.com/abstract=539843.

Lotz S (2015) Spontaneous giving under structural inequality: Intuition promotes cooperation in asymmetric social dilemmas. *PLoS ONE* 10(7):e0131562.

Mailath GJ, Samuelson L (2006) *Repeated Games and Reputations: Long-Run Relationships* (Oxford University Press, Oxford, UK).

McAllister DJ (1995) Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Acad. Management J.* 38(1):24–59.

Nunn N, Wantchekon L (2011) The slave trade and the origins of mistrust in Africa. *Amer. Econom. Rev.* 101(7):3221–3253.

Ostrom E, Walker J, Gardner R (1992) Covenants with and without a sword: Self-governance is possible. *Amer. Political Sci. Rev.* 86(2):404–417.

Ouss A, Peysakhovich A (2015) When punishment doesn't pay: Cold glow and decisions to punish. *J. Law Econom.* 58(3): 625–655.

Pedersen E, Kurzban R, McCullough M (2013) Do humans *really* punish altruistically? A closer look. *Proc. Roy. Soc. Ser. B: Biol. Sci.* 280(1758):Article 20122723.

Peysakhovich A, Plagborg-Møller M (2012) A note on proper scoring rules and risk aversion. *Econom. Lett.* 117(1):357–361.

Peysakhovich A, Nowak MA, Rand DG (2014) Humans display a "cooperative phenotype" that is domain general and temporally stable. *Nature Comm.* 5(September):Article 4939.

Putnam RD (2000) *Bowling Alone: The Collapse and Revival of American Community* (Simon & Schuster, New York).

Rabin M (1993) Incorporating fairness into game theory and economics. *Amer. Econom. Rev.* 83(5):1281–1302.

Rand DG, Nowak MA (2011) The evolution of antisocial punishment in optional public goods games. *Nature Comm.* 2(August): Article 434.

Rand DG, Nowak MA (2012) Evolutionary dynamics in finite populations can explain the full range of cooperative behaviors observed in the centipede game. *J. Theoret. Biol.* 300:212–221.

Rand DG, Nowak MA (2013) Human cooperation. *Trends Cognitive Sci.* 17(8):413–425.

Rand DG, Fudenberg D, Dreber A (2015a) It's the thought that counts: The role of intentions in noisy repeated games. *J. Econom. Behav. Organ.* 116(August):481–499.

Rand DG, Greene JD, Nowak MA (2012) Spontaneous giving and calculated greed. *Nature* 489(7416):427–430.

Rand DG, Newman GE, Wurzbacher O (2015b) Social context and the dynamics of cooperative choice. *J. Behavioral Decision Making* 28(2):159–166.

Rand DG, Yoeli E, Hoffman M (2014a) Harnessing reciprocity to promote cooperation and the provisioning of public goods. *Policy Insights Behavioral Brain Sci.* 1(1):263–269.

Rand DG, Tarnita CE, Ohtsuki H, Nowak MA (2013) Evolution of fairness in the one-shot anonymous ultimatum game. *Proc. Natl. Acad. Sci. USA* 110(7):2581–2586.

Rand DG, Dreber A, Ellingsen T, Fudenberg D, Nowak MA (2009) Positive interactions promote public cooperation. *Science* 325(5945):1272–1275.

Rand DG, Peysakhovich A, Kraft-Todd GT, Newman GE, Wurzbacher O, Nowak MA, Green JD (2014b) Social heuristics shape intuitive cooperation. *Nature Comm.* 5(April): Article 3677.

Richerson PJ, Boyd R (2005) *Not by Genes Alone: How Culture Transformed Human Evolution* (University of Chicago Press, Chicago).

Roch SG, Lane JAS, Samuelson CD, Allison ST, Dent JL (2000) Cognitive load and the equality heuristic: A two-stage model of resource overconsumption in small groups. *Organ. Behav. Human Decision Processes* 83(2):185–212.

Roth AE (2002) The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica* 70(4):1341–1378.

Rousseau DM, Sitkin SB, Burt RS, Camerer C (1998) Not so different after all: A cross-discipline view of trust. *Acad. Management Rev.* 23(3):393–404.

Samuelson L (2001) Introduction to the evolution of preferences. *J. Econom. Theory* 97(2):225–230.

Sapienza P, Zingales L, Guiso L (2006) *Does Culture Affect Economic Outcomes?* (National Bureau of Economic Research, Cambridge, MA).

Schulz JF, Fischbacher U, Thöni C, Utikal V (2014) Affect and fairness: Dictator games under cognitive load. *J. Econom. Psych.* 41(April):77–87.

Shenhav A, Rand DG, Greene JD (2012) Divine intuition: Cognitive style influences belief in God. *J. Experiment. Psych.: General* 141(3):423–428.

Simmons JP, Nelson LD, Simonsohn U (2013) Life after P-hacking. *Meeting Soc. Personality Soc. Psych., New Orleans*. http://ssrn.com/abstract=2205186.

Tabellini G (2008) The scope of cooperation: Values and incentives. *Quart. J. Econom.* 123(3):905–950.

Thaler RH, Benartzi S (2004) Save More Tomorrow™: Using behavioral economics to increase employee saving. *J. Political Econom.* 112(S1):S164–S187.

Thaler RH, Sunstein CR (2008) *Nudge: Improving Decisions About Health, Wealth, and Happiness* (Yale University Press, New Haven, CT).

Tinghög G, Andersson D, Bonn C, Böttiger H, Josephson C, Lundgren G, Västfjäll D, Kirchler M, Johannesson M (2013) Intuition and cooperation reconsidered. *Nature* 497(7452):E1–E2.

Tversky A, Kahneman D (1974) Judgment under uncertainty: Heuristics and biases. *Science* 185(4157):1124–1131.

Verkoeijen PPJL, Bouwmeester S (2014) Does intuition cause cooperation? *PLoS ONE* 9(5):e96654.

Welsh MB, Burns NR, Delfabbro PH (2013) The cognitive reflection test: How much more than numerical ability? Knauff M, Sebanz N, Pauen M, Wachsmuth I, eds. *Proc. 35th Annual Meeting Cognitive Sci. Soc.* (Cognitive Science Society, Austin, TX), 1587–1592.

Yamagishi T, Horita Y, Mifune N, Hashimoto H, Li Y, Shinada M, Miura A, Inukai K, Takagishi H, Simunovic D (2012) Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity. *Proc. Natl. Acad. Sci. USA* 109(50): 20364–20368.

Yoeli E, Hoffman M, Rand DG, Nowak MA (2013) Powering up with indirect reciprocity in a large-scale field experiment. *Proc. Natl. Acad. Sci. USA* 110(Supplement 2):10424–10429.