

ULRR

Maintaining cooperation through vertical communication of trust when removing sanctions

Item Type	Article
Authors	Posten, Ann-Christin;Uğurlar, Pinar;Kube, Sebastian;Lammers, Joris
Citation	Proceedings of The National Academy of Sciences of the United States of America, 122, 12, e2415010122
DOI	10.1073/pnas.2415010122
Publisher	PNAS
Rights	Attribution-NonCommercial-ShareAlike 4.0 International
Download date	2026-05-11 23:52:47
Item License	http://creativecommons.org/licenses/by-nc-sa/4.0/
Link to Item	https://hdl.handle.net/10344/31446



Maintaining cooperation through vertical communication of trust when removing sanctions

Ann-Christin Posten^{a,1} , Pinar Uğurlar^b , Sebastian Kube^{c,d,e}, and Joris Lammers^f

Affiliations are included on p. 5.

Edited by Karen Cook, Stanford University, Stanford, CA; received August 28, 2024; accepted January 29, 2025

An effective way to foster cooperation is to monitor behavior and sanction freeriding. Yet, previous studies have shown that cooperation quickly declines when sanctioning mechanisms are removed. We test whether explicitly expressing trust in players' capability to maintain cooperation after the removal of sanctions, i.e., vertical communication of trust, has the potential to alleviate this drop in compliance. Four incentivized public-goods experiments ($N = 2,823$) find that the vertical communication of trust maintains cooperation upon the removal of centralized (Study 1), third-party (Study 2a, 2b), and peer punishment (Study 3), and this effect extends beyond single interactions (Study 4). In all studies, vertical trust communication increases mutual trust among players, providing support to the idea that vertically communicating trust can be a self-fulfilling prophecy. Extrapolating our findings to natural environments, they suggest that authorities should carefully consider how they communicate the lifting of rules and sanctions.

cooperation | vertical trust | punishment | public good | experiment

Modern social and environmental challenges are often rooted in cooperation problems. These cooperation problems emerge due to a conflict between individual rationality and the collective perspective (1). Among the most popular mechanisms to mitigate such cooperation problems is monitoring with various forms of sanctioning institutions. For example, in response to the COVID-19 pandemic, many countries introduced sanctions to enforce compliance with cooperative public health measures. Yet, when the need for these measures decreased—because of dropping infection rates and increasing vaccination rates—political pressure put an end to the sanctioning mechanisms and selfish behavior increased (2). Although it is in the collective interest to voluntarily continue to cooperate, this detrimental effect on cooperation rates after the removal of sanctions has been frequently shown in lab studies (3–6). This underscores the importance of finding ways to sustain cooperation, even after the removal of sanctioning mechanisms. Maintaining cooperation after the removal of sanctions can involve emphasizing trust, shared responsibility, and the benefits of collective action. Denmark provides an illustrative case of addressing this challenge. The government lifted most COVID-19 restrictions in 2021, citing trust in citizens to act responsibly and high vaccination rates. Rather than relying on continued enforcement, Danish Prime Minister Mette Frederiksen encouraged citizens to independently assess their needs for precautions. This trust-based approach sought to maintain public cooperation and collective responsibility even in the absence of formal sanctioning mechanisms, demonstrating a governance approach based on trust, cooperation, and shared responsibility (7, 8). In this paper, we thus ask whether communication from authorities can play a critical role in sustaining cooperation after the removal of sanctioning mechanisms—not only in the context of centralized punishment (as in the COVID-19 example) but also more generally for punishment by a (otherwise unrelated) third-party and mutual second-party (peer) punishment. Our specific focus lies on the (unidirectional) vertical communication from the principal, the authority responsible for removing the punishment mechanism, to the agents encountering the cooperation problem. More precisely, we empirically test whether cooperation can be maintained if the relevant authorities explain—when they communicate the removal of the punishment institution—that they do so because they trust the population to cooperate in its absence.

We test this question by changing the experimenters' expression of trust in participants' cooperativeness, subsequent to the removal of a sanctioning institution.* This choice follows the notion that participants in experiments often look to experimenters as authority figures,

Significance

An effective way to ensure cooperation is to sanction noncompliance. Yet, once sanctions are removed, cooperation tends to drop. We demonstrate that vertical communication by authorities can help maintain cooperation when abolishing punishment institutions. Specifically, we show that when authorities communicate their trust in the population to cooperate in the absence of punishment institutions, this acts as a self-fulfilling prophecy and leads people to maintain cooperation even in the absence of punishment. Our findings complement previous literature on (horizontal) norms, beliefs, and conditional cooperation. They offer valuable insights that can help authorities maintain voluntary cooperation, even after enforcement mechanisms are terminated (for example, after a pandemic) or extremely difficult to monitor (for example, energy conservation regulations).

Author contributions: A.-C.P., P.U., S.K., and J.L. designed research; A.-C.P. performed research; A.-C.P. analyzed data; and A.-C.P., P.U., S.K., and J.L. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2025 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹To whom correspondence may be addressed. Email: ann-christin.posten@ul.ie.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2415010122/-/DCSupplemental>.

Published March 17, 2025.

*In our four main experiments, the communicating authority is the experimenter, and in a complementary experiment (Study 2b), it is a participant who is granted this authority.

much like citizens look to the government (9). Although existing literature has often studied the negative consequences of this, such as in Milgram's (1974) classic studies where participants' obedience to the authority of experimenters produced cruelty and immorality (10), here we instead study a potential positive consequence: Does the vertical expression of trust by experimenters in participants' ability to cooperate without sanctioning institutions promote cooperation?

By investigating the efficacy of *vertical* communication (from the principal to the population), we complement the prevalent focus in the existing literature on horizontal communication (among the members of the population) as a substitute for punishment to foster cooperation (11, 12). We expect vertical communication to be an effective instrument for maintaining cooperation when sanctioning mechanisms are removed, because providing the explanation of trust driving the change might act as a self-fulfilling prophecy (13): The vertical communication of trusting the population to cooperate shapes the beliefs within the population that others can still be trusted, which in turn sustains their willingness to cooperate, thus maintaining cooperation levels and showing that others can indeed be trusted to cooperate.

This line of reasoning is similar to an equilibrium-selection idea for societies harboring multiple equilibria. The idea is proposed by Dasgupta (14), for example, who also stresses the importance of mutual trust as a basis for cooperation and Cook et al., who argue that trust can create microlevel environments that lower the costs of monitoring and sanctioning (15). Indeed, Thöni et al. (6) observe a positive correlation between trust, as measured by standard trust questions from the World Values Survey, and cooperation behavior in an incentivized cooperation game. Likewise, Gill and Rosokha (16) find that more trusting subjects, measured by a personality survey, cooperate more in an indefinitely repeated prisoner's dilemma. This link is also found by Kim et al. (17) when combining lab experimental data from trust and cooperation games. They show that people infer cooperativeness from trustworthiness and cooperate more when they predict others to be more cooperative. Similarly, based on survey data, Schmelz (18) stresses the essential role of trust in government for citizens' voluntary and nonenforced cooperation with COVID-19 policies. Complementing these behavioral findings on trust and cooperation, simulation results reported in Battu and Rahwan (19) suggest that "society should focus on creating a critical amount of trust to harness the conditional nature of its members."

Vertical communication is of particular significance, given the importance of conditionality and beliefs about others cooperativeness for shaping subsequent cooperative behavior. If a sizable fraction of a population acts like conditional cooperators (20), the beliefs that people hold about one another become key to cooperation (21, 22), and shaping beliefs through vertical communication should have significant effects on subsequent behavior. Consistent with this idea, Engel et al. (23) find that groups cooperate more after being provided with selective positive information about the cooperative behavior of other, unrelated groups from the same population. Vice versa, Galbiati et al. (24) observe that sanctions can be perceived as a negative signal about others' cooperativeness, if they are actively installed by an informed third party, and thus make cooperation in a minimum-effort game more difficult. Relatedly, but focusing on horizontal communication, Tyran and Feld (25), Gülerk et al. (5), and Sutter et al. (26) all demonstrate that endogenous institutional choices, made by the population via voting, can serve as a signal of others' cooperativeness.

Results

To provide causal evidence for the effectiveness of vertical communication on cooperation behavior, we preregistered and conducted four incentivized main experiments (Studies 1, 2a, 3, and 4) and one complementary experiment (Study 2b), testing the effectiveness of the vertical communication of trust.[†] All studies employed an incentivized public-goods game (PGG), a workhorse frequently used in the literature to study cooperation problems. In PGGs, participants form groups and individually decide how many tokens from their private endowment they want to contribute to a common pool. This pool is then multiplied and shared equally among all group members. The game parameters are chosen such that the group payoff is maximized if all tokens are contributed, but selfish money-maximizing behavior leads to the inefficient equilibrium outcome of zero contributions. At the beginning, all participants were informed about the total number of rounds they were about to play, and sanctioning parameters were explained. Players did not receive any feedback on payoffs before the very end of the study. The setup was completely anonymous, and the groups were randomly drawn each round. Each study consisted of a first round, where the PGG was accompanied by a sanctioning institution, followed by a second round, where the sanctions were removed. The three sanctioning institutions that we used are the most prominent ones in the literature on cooperation and punishment and closely followed frequently used parameterizations. Study 1 used a centralized punishment scheme, Study 2a/b a third-party punishment scheme, and Studies 3 and 4 a peer punishment scheme. The first round was followed by a second round of the PGG without any sanctioning institution (except for Study 4, which included a punishment-maintaining control condition). In each study, treatment conditions varied by how the removal of sanctions between the first and second rounds was communicated to the participants (i.e., trust communication vs. distrust communication vs. no trust-related communication). In addition to using contributions as a proxy for cooperation behavior, trust was assessed by using a standard Likert-type question. For all three sanctioning institutions, we found that the vertical communication of trust affects self-reported trust and helps to maintain cooperation despite the removal of the sanctioning mechanism. Upon vertical trust communication, across all main studies, contributions were on average 25% higher than those without communicating a reason for sanctioning removal.

Study 1 ($n = 604$) investigated whether the vertical communication of trust can maintain cooperation after a *centralized* sanctioning mechanism is removed. A centralized mechanism closely mirrors sanctioning mechanisms imposed by authorities, similar to our opening example on COVID-19. The obligation under this mechanism was to invest the socially optimal amount into the public investment. If a player was detected to deviate from this obligation, that person had to pay a nondeterrent fine depending on the size of the deviation. This was implemented as follows: One participant of each group was randomly drawn and punished via the subtraction of 1.6 tokens for each token that this player had not invested in the collective [see, for instance, Galbiati and Vertova (27) for a similar approach]. In the second round, the sanctioning mechanism was removed, and the vertical communication took place. When introducing the removal, participants were informed that this happened because the researchers either trusted them (*trust condition*), did not trust them (*distrust condition*), or did not receive any explanation (*control condition*). Specifically, in the trust condition, the experimenters stated, "We trust that most people invest sufficiently," whereas in the distrust condition, they stated, "We do not trust that

[†]Studies were conducted in a different chronological order than the order in which they appear here. We thank two anonymous referees for suggesting complementary Study 2b.

most people invest sufficiently.” After their decision in round 2, participants indicated their trust in their coplayers and additional measures (see *SI Appendix* for further study descriptions and analyses, particularly *SI Appendix*, Table S1 for expectations on trust and cooperation and *SI Appendix*, Table S2 for affective states).

Our main variable of interest was the contributed amount as the proxy for cooperation. Results showed that vertical communication of trust kept contributions at a higher level upon sanctioning removal compared to the control and distrust condition (see Fig. 1A for unadjusted means analysis), while contributions in the control and distrust condition dropped more strongly in their absence, interaction effect $F(2, 601) = 3.26, P = 0.039, \eta^2 = 0.011$. This resulted in significantly higher contributions in the trust condition as compared to the distrust ($P = 0.007$) and control condition ($P = 0.010$), with the latter two not differing significantly ($P = 0.920$).[‡] This pattern supports the prediction that people by default expect others to defect upon removal of a sanctioning mechanism and that therefore cooperation drops—but that vertical communication of trust in the ability to contribute even in the absence of sanctions mitigates that drop in cooperation. In fact, shedding more light on the underlying psychological process by a multicategorical mediation analysis, with round 1 contribution as covariate, suggests that trust in coplayers, assessed by the seven-point Likert-type question “Do you think the other players could be trusted?,” acted as a critical psychological process-variable, mediating the effect of trust and distrust on cooperation (Fig. 1A; see *SI Appendix* for full mediation analysis results).

Study 2a ($n = 606$) investigated whether the vertical communication of trust can maintain cooperation after a *third-party* sanctioning mechanism is removed. This is comparable to judicial systems across the world where court judges have some discretion over the punishment [see, for instance, Fehr and Fischbacher (28)]. Different from the centralized institution in Study 1, where

obligation and fines were known in advance and carried out automatically, both the basis for and the size of the punishment were now decided by an actual player outside the group. Therefore, at the very end of the study, one randomly selected player of each group was given the opportunity to punish the players of another group by subtracting up to 15 tokens (which were destroyed) from their individual payoffs. The third party was able to tailor the size of the punishment to the contribution levels, using a strategy method approach.

After the first round, the punishment mechanism was revoked with the same vertical communication as before, again creating a trust, control, and distrust condition. As in Study 1, an interaction effect confirmed, $F(2, 603) = 13.52, P < 0.001, \eta^2 = 0.04$, that, upon removal of the sanctioning mechanism, vertically communicating trust in people affected contributions. Planned contrasts showed that vertical trust communication kept contributions at a significantly higher level compared to when communicating distrust ($P < 0.001$), or without communication in the control condition ($P < 0.001$). No difference between the distrust and control conditions emerged ($P = 0.906$). As in Study 1, trust in others acted as a mediating process variable (Fig. 1B; see *SI Appendix* for full mediation analysis results).

So far, it has always been the experimenter, as the authority figure, that communicated trust. While this is a natural choice in centralized punishment (Study 1) and peer-punishment (Study 3, discussed below), a complementary approach to Study 2a could involve the third-party punisher communicating trust while removing the third-party punishment institution. In complementary Study 2b ($n = 400$), we conducted both a trust and control condition, to examine whether the effects of vertical trust communication can also be observed when the decision to remove punishment in round 2 (or not) stems from the participant who is in the role of the third-party punisher. Similar to Study 2a, in Study 2b, an interaction effect confirmed that vertical trust

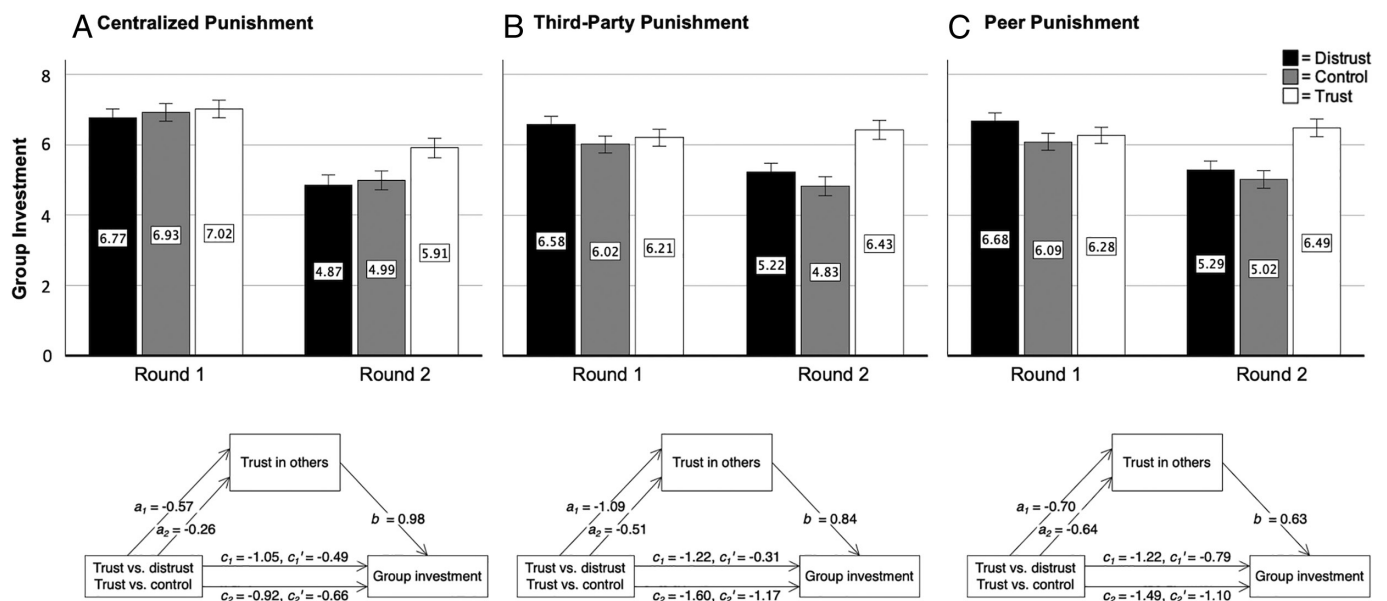


Fig. 1. Mean contributions for Studies 1, 2a, and 3. Upon removal of sanctioning mechanisms in round 2, the vertical communication of trust as reason for the removal leads to more cooperation (group investment), measured as tokens contributed to a public good, than no communication (control condition) or communicating distrust (distrust condition). This was true for centralized punishment in Study 1 (A), third-party punishment in Study 2a (B), and peer punishment in Study 3 (C). Trust in others mediated the relationship between vertical trust communication and group investment comparing the trust with the distrust condition (path a_1) in all three studies and with the control condition (path a_2) in Studies 2a and 3. Conditions in the multicategorical mediation analysis were dummy-coded as 0 for trust, 1 for distrust, and 2 for control. Error bars represent ± 1 SE.

[‡]For all studies, planned contrasts controlled for participants’ individual propensity to cooperate (i.e., round 1 contributions).

communication affected contributions in round 2 compared to round 1, $F(1, 398) = 19.64$, $P < 0.001$, with significant differences between the trust condition and control condition ($P < 0.001$). These results indicate that also authorities beyond the experimenters themselves can establish cooperation among the people they are responsible for by combining a vertical communication of trust with the removal of sanctions.

Aside from formal institutions and their agents, also fellow citizens and peers can punish noncooperative behavior. Study 3 ($n = 610$) mimicked this by employing a *peer* punishment mechanism, as it is frequently used in existing literature (4). This time, one randomly drawn player per group had the opportunity to punish the members of the same group. Again, punishment decisions were provided at the end of the study, via the strategy method. Results showed that, upon removal of the sanctions, contributions were again affected by the vertical communication of trust, interaction effect, $F(2, 607) = 14.37$, $P < 0.001$, $\eta^2 = 0.05$, resulting in significantly lower contributions in the distrust ($P < 0.001$) and control condition ($P < 0.001$), compared to the trust condition. Again, there was no difference between the distrust and control condition ($P = 0.741$). Trust in others again acted as a mediating process variable (Fig. 1C; see *SI Appendix* for full mediation analysis results).

In Studies 1 to 3, we always compared the effect of vertical trust communication (vs. distrust or control) while removing a sanction on contributions in a second round. Study 4 ($n = 603$) addressed two potential limitations. First, it directly compared the effect of vertical communication of trust with that of maintaining a sanctioning mechanism. Second, it tested whether the effect of vertical communication of trust maintains over time. Therefore, we replicated Study 3 and extended the design to three rounds of PGGs. In the first round, the peer punishment mechanism was in place for all participants. In round 2, we compared three conditions: We used the same trust and distrust manipulations as before (conditions 1 and 2), but this time also included a third condition, in which the sanctioning mechanism was maintained in round 2 (i.e., nothing changed compared to round 1). This enabled us to examine how effective the vertical trust communication is compared to maintaining the sanctioning institution. In the third round, the sanctioning mechanism was removed in the third condition, without providing a reason for the removal. In the trust and distrust conditions, the sanctioning institution remained absent in the third round. Thus, all participants played the public goods game in the third round without a sanctioning institution being in place. Inspecting cooperation in round 2 showed that the trust condition was comparably high as the third condition

with the ongoing sanctioning mechanism and did not differ significantly ($P = 0.558$), suggesting that vertical communication of trust when removing sanctions is similarly effective as maintaining the sanctioning mechanism (Fig. 2). The distrust condition dropped in comparison to both the trust condition ($P < 0.001$) and the maintained-sanctioning condition ($P < 0.001$). Results of round 3 (in which none of the conditions featured a sanctioning mechanism) showed the clear advantage of the previous vertical communication of trust when removing sanctioning: Although the round 3 instructions were identical across all three conditions, the mere prior vertical communication of trust in round 2 ensured that contributions in round 3 were higher in the trust condition than in the distrust condition ($P = 0.004$) and the (third) control condition ($P = 0.004$). Distrust and control conditions did not differ ($P = 0.989$), highest-order interaction, $F(4, 1,200) = 12.84$, $P < 0.001$, $\eta^2 = 0.041$.

Discussion

Sanctioning institutions have been shown to effectively foster cooperation (29). However, reasons might arise that lead to the institutions being removed, such as diminishing political support or increasing implementation cost. In fact, both Markussen et al. (30) and Kamei et al. (31) demonstrate that, in the case of a cooperation problem, as the cost of a sanctioning institution increases, participants tend to vote for the implementation of alternative measures that are less costly. Also, the alternative of keeping up the obligations without backing it up by sanctions for misbehavior is usually only of limited efficacy (27). Yet, maintaining cooperation in the absence of sanctioning institutions is of great societal importance. In this paper, we explored whether the vertical communication of trust from authorities to the population can maintain cooperation. For three popular sanctioning institutions, centralized punishment, peer punishment, and third-party punishment, we found that one can maintain cooperation, despite the removal of sanctions, by communicating trust in people's continued cooperation (Studies 1 to 3). This act of communication served as a self-fulfilling prophecy, enabling cooperation to be sustained at a level similar to when sanctions were still in place (Study 4). Additionally, the positive consequences of this top-down trust-based approach persisted beyond one immediate decision to contribute, whereas the effects of sanctions vanished immediately in their absence (Study 4). Notably, the removal of sanctioning mechanism without any explanation led to an immediate drop in cooperation and trust in others, similar to actively communicating distrust (Studies 1, 2a, 3, and 4), suggesting that

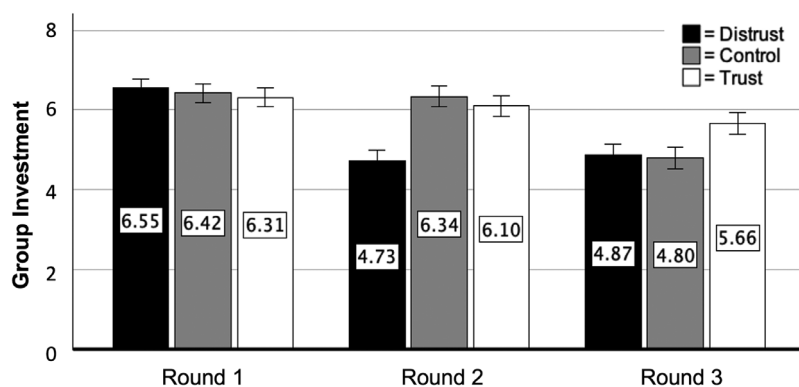


Fig. 2. Mean contributions in Study 4. Removing sanctioning mechanisms in round 2 together with vertical trust communication (trust condition) led to no significant difference in contributions to the public good compared to the control condition, where sanctions were still installed, and to higher contributions than if distrust was communicated (distrust condition). In round 3, when punishment institutions were absent in all three conditions, having signaled trust prior to round 2 yielded higher contributions in round 3 compared to the other two conditions in which no trust had been communicated. Error bars represent ± 1 SE.

the presence of sanctioning mechanisms creates an expectation that others are not to be trusted and not willing to cooperate in their absence.

We measured mood and other affective states (32) in relation to the last round of the PGG, in which our critical differences between trust, control, and distrust condition occurred. In all four studies, results suggested that mood did not account for the observed effects because the pattern of results did not mirror the pattern of cooperative behavior (*SI Appendix, Table S2*). This is consistent with other work suggesting that mood does not have strong direct effects on cooperative behavior (33).

Our findings add to the central importance of trust in cooperation (34). The removal of sanctions undermines any potential default trust that people may have held before (35–37). It leads to the quick inference that others may be untrustworthy and cause a self-fulfilling prophecy of defection. We show that a brief vertical communication of trust prevents this and instills a self-fulfilling prophecy of trust that maintains compliance even in the absence of enforcement. These insights are also interesting for researchers working on conditional cooperation, because they add substantially to established findings on the importance of beliefs about others' cooperation and demonstrate an alternative method for shaping these beliefs. Moreover, our mediation analyses suggest that the same psychological process-variable, self-reported trust in other group members, is critical in all three sanctioning institutions. Furthermore, the findings tap into a gap in research on the consequences of trust on its recipients (38), by showing that vertical trust communication can create horizontal trust within the population of trustees. These insights are interesting for practitioners, because they identify vertical communication as an easy and effective way of maintaining cooperation. In particular, when cooperative behavior has developed over a longer timespan under a sanctioning institution that is about to be abolished, communicating trust and confidence in people's ability to act in a considerate manner not only appears quite powerful but also cost-effective.

Follow-up research could examine how vertical communication of trust compares to the communication of different content, different forms of communication [e.g., via behavioral expression (39)], or by different sources (e.g., senders with or without stakes in the underlying cooperation game or in-group compared to out-group senders). The communication of injunctive or descriptive norms might be of particular relevance (40, 41), since the authority's vertical trust communication in our study might also have shifted the focus to norms of cooperation (42, 43). In future research, the comparison of our results to conditions where norms are explicitly communicated would allow to disentangle the focusing and informational effect of norms (44) from the effect of inducing trust and shaping beliefs for conditional cooperators. Measuring norms explicitly could also contribute to our understanding whether removal of sanctioning shifts the perceived relational model and its corresponding norms on cooperation, as suggested by Relational Models Theory (45). Moreover, it would be interesting to see in future research whether the beneficial effects of vertical trust communication extend to other settings in which cooperation is critical (e.g. organizations or educational institutions), particularly

costly (e.g., when persistent behavioral change is required), or extremely difficult to monitor [e.g., energy conservation regulations (46)].

Materials and Methods

Participants could only take part in one of the studies. Informed consent was obtained from all participants. The studies were run online using Qualtrics in 2021 to 2024. Participants were US citizens who were randomly sampled from the participant pool of Prolific, a UK-based service provider specialized in conducting academic research. Payments on Prolific were at that time generally made in Pounds Sterling (GBP), so we used British Pence throughout our studies, paying GBP 0.10 per 10 tokens earned in the game (in addition to a base payment). Studies were conducted consistent with ethical principles provided by the German Psychological Society (47) and reviewed by the Ethics Committee of the Faculty of Human Sciences of the University of Cologne. Data, code, and materials are deposited through the Open Science Framework (OSF) at <https://doi.org/10.17605/osf.io/duvrc> (48). Studies were preregistered at <https://aspredicted.org/> and are accessible via OSF (Studies 1, 2a, 2b, 3, 4) (48).

The basic underlying game in each study was a four-player PGG, a game that is commonly used in the literature to study cooperation. Each round, players decided how to distribute 10 tokens between an individual and a collective investment. All tokens invested individually were tripled and returned only to the investing player, while all tokens invested collectively were doubled, and the doubled amount was paid out to each of the four players in the group [following a similar parameterization as used, for instance, in Andreoni and Gee (49)]. The individual payoff-maximizing strategy is to keep own tokens in the individual investment pot, independent of what the others (are expected to) do, in which case each player would earn $10 \times 3 = 30$ tokens from the respective private investment. The cooperative outcome is to invest the entire endowment in the collective pot, in which case each player would earn a return of $(4 \times 10) \times 2 = 80$ tokens from the public investments. The maximum payoff per round for an individual player was 90 tokens if the player only invested individually and the other three players only invested in the collective (i.e., 10×3 tokens from the private investment and $(3 \times 10) \times 2 = 60$ tokens from the other three players' public investments). The minimum payoff was 20 tokens if the player only invested in the public good and all other players only invested individually (i.e., $10 \times 2 = 20$ tokens from the public investment and $(3 \times 0) \times 2 = 0$ tokens from the other three players' missing public investment).

Data, Materials, and Software Availability. Data, code, and materials are deposited through the OSF at <https://doi.org/10.17605/osf.io/duvrc> (48). Studies were preregistered at <https://aspredicted.org/> and are accessible via OSF (Studies 1, 2a, 2b, 3, and 4) (48).

ACKNOWLEDGMENTS. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC 2126/1-390838866 and the research grant PO 1850/3-1 awarded to A.-C.P. Declaration of AI and AI-assisted technologies in the writing process: During the preparation of this work, we used you.com in order to improve readability and language. After using this service, we reviewed and edited the content as needed and take full responsibility for the content of the publication.

Author affiliations: ^aDepartment of Psychology, University of Limerick, Limerick V94 T9PX, Ireland; ^bDepartment of Psychology, Özyeğin University, Istanbul 34794, Türkiye; ^cInstitute for Applied Microeconomics, University of Bonn, Bonn 53113, Germany; ^dCenter for Economics and Neuroscience, University of Bonn, Bonn 53113, Germany; ^eMax Planck Institute for Research on Collective Goods, Research Group Behavioral Law and Economics, Bonn 53113, Germany; and ^fDepartment of Psychology, Social Cognition Center Cologne, University of Cologne, Cologne 50931, Germany

1. E. Fehr, I. Schurtenberger, Normative foundations of human cooperation. *Nat. Hum. Behav.* **2**, 458–468 (2018).
2. J. C. Yong, B. K. C. Choy, Noncompliance with safety guidelines as a free-riding strategy: An evolutionary game-theoretic approach to cooperation during the COVID-19 pandemic. *Front. Psychol.* **12**, 646892 (2021).
3. E. Fehr, S. Gächter, Cooperation and punishment in public goods experiments. *Am. Econ. Rev.* **90**, 980–994 (2000).

4. E. Fehr, S. Gächter, Altruistic punishment in humans. *Nature* **450**, 137–140 (2002).
5. Ö. Gürekr, B. Irlenbusch, B. Rockenbach, The competitive advantage of sanctioning institutions. *Science* **312**, 108–111 (2006).
6. B. Herrmann, C. Thöni, S. Gächter, Antisocial punishment across societies. *Science* **319**, 1362–1367 (2008).
7. L. Klingert, Denmark is first in EU to lift all coronavirus restrictions. *Politico* (2021). <https://www.politico.eu/article/denmark-first-eu-lift-coronavirus-restrictions/>. Accessed 30 November 2024.
8. M. B. Petersen, COVID lesson: Trust the public with hard truths. *Nature* **598**, 237–237 (2021).

9. R. Spears, H. J. Smith, Experiments as politics. *Polit. Psychol.* **22**, 309–330 (2001).
10. S. Milgram, *Obedience to Authority: An Experimental View* (Harper Row, 1974).
11. O. Bochet, T. Page, L. Putterman, Communication and punishment in voluntary contribution experiments. *J. Econ. Behav. Organ.* **60**, 11–26 (2006).
12. J. C. Cardenas, J. Stranlund, C. Willis, Local environmental control and institutional crowding-out. *World Dev.* **28**, 1719–1733 (2000).
13. M. Bacharach, G. Guerra, D. J. Zizzo, The self-fulfilling property of trust: An experimental study. *Theory Decis.* **63**, 349–388 (2007).
14. P. Dasgupta, Trust and cooperation among economic agents. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364**, 3301–3309 (2009).
15. K. S. Cook, R. Hardin, M. Levi, *Cooperation without Trust?* (Russell Sage Foundation, 2005).
16. D. Gill, Y. Rosokha, Beliefs, learning, and personality in the indefinitely repeated prisoner's dilemma. *Am. Econ. J. Microecon.* **16**, 259–283 (2024).
17. J. Kim, L. Putterman, X. Zhang, Trust, beliefs and cooperation: Excavating a foundation of strong economies. *Eur. Econ. Rev.* **147**, 104166 (2022).
18. K. Schmelz, Enforcement may crowd out voluntary support for COVID-19 policies, especially where trust in government is weak and in a liberal society. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2016385118 (2021).
19. B. Battu, T. Rahwan, Cooperation without punishment. *Sci. Rep.* **13**, 1213 (2023).
20. U. Fischbacher, S. Gächter, E. Fehr, Are people conditionally cooperative? Evidence from a public goods experiment. *Econ. Lett.* **71**, 397–404 (2001).
21. U. Fischbacher, S. Gächter, Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *Am. Econ. Rev.* **100**, 541–556 (2010).
22. S. Gächter, F. Kölle, S. Quercia, Reciprocity and the tragedies of maintaining and providing the commons. *Nat. Hum. Behav.* **1**, 650–656 (2017).
23. C. Engel, S. Kube, M. Kurschilgen, Managing expectations: How selective information affects cooperation and punishment in social dilemma games. *J. Econ. Behav. Organ.* **187**, 111–136 (2021).
24. R. Galbiati, K. H. Schlag, J. J. Van Der Weele, Sanctions that signal: An experiment. *J. Econ. Behav. Organ.* **94**, 34–51 (2013).
25. J. Tyran, L. P. Feld, Achieving compliance when legal sanctions are non-deterrent. *Scand. J. Econ.* **108**, 135–156 (2006).
26. M. Sutter, S. Haigner, M. G. Kocher, Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations. *Rev. Econ. Stud.* **77**, 1540–1566 (2010).
27. R. Galbiati, P. Vertova, Obligations and cooperative behaviour in public good games. *Games Econ. Behav.* **64**, 146–170 (2008).
28. E. Fehr, U. Fischbacher, Third-party punishment and social norms. *Evol. Hum. Behav.* **25**, 63–87 (2004).
29. D. Balliet, L. B. Mulder, P. A. M. Van Lange, Reward, punishment, and cooperation: A meta-analysis. *Psychol. Bull.* **137**, 594–615 (2011).
30. T. Markusen, L. Putterman, J.-R. Tyran, Self-organization for collective action: An experimental study of voting on sanction regimes. *Rev. Econ. Stud.* **81**, 301–324 (2014).
31. K. Kamei, L. Putterman, J.-R. Tyran, State or nature? Endogenous formal versus informal sanctions in the voluntary provision of public goods. *Exp. Econ.* **18**, 38–65 (2015).
32. N. Schwarz, G. L. Clore, Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *J. Pers. Soc. Psychol.* **45**, 513–523 (1983).
33. G. Hertel, J. Neuhof, T. Theuer, N. L. Kerr, Mood effects on cooperation in small groups: Does positive mood simply lead to more cooperation? *Cogn. Emot.* **14**, 441–472 (2000).
34. D. Balliet, P. A. M. Van Lange, Trust, conflict, and cooperation: A meta-analysis. *Psychol. Bull.* **139**, 1090–1112 (2013).
35. Y. Schul, R. Mayo, E. Burnstein, Encoding under trust and distrust: The spontaneous activation of incongruent cognitions. *J. Pers. Soc. Psychol.* **86**, 668–679 (2004).
36. Y. Schul, R. Mayo, E. Burnstein, The value of distrust. *J. Exp. Soc. Psychol.* **44**, 1293–1302 (2008).
37. M. Katzir, A.-C. Posten, Are there dominant response tendencies for social reactions? Trust trumps mistrust—Evidence from a Dominant Behavior Measure (DBM). *J. Pers. Soc. Psychol.* **125**, 57–81 (2023).
38. O. Schilke, M. Reimann, K. S. Cook, Trust in social relations. *Annu. Rev. Sociol.* **47**, 239–259 (2021).
39. G. Spadaro *et al.*, Corrupt third parties undermine trust and prosocial behaviour between people. *Nat. Hum. Behav.* **7**, 46–54 (2022).
40. R. B. Cialdini, R. R. Reno, C. A. Kallgren, A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *J. Pers. Soc. Psychol.* **58**, 1015 (1990).
41. K. Keizer, S. Lindenberg, L. Steg, The importance of demonstratively restoring order. *PLoS One* **8**, e65137 (2013).
42. S. Gavrilets, P. J. Richerson, Collective action and the evolution of social norm internalization. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 6068–6073 (2017).
43. M. M. Blair, L. A. Stout, Trust, trustworthiness, and the behavioral foundations of corporate law. *Univ. Pa. Law Rev.* **149**, 1735 (2001).
44. E. Krupka, R. A. Weber, The focusing and informational effects of norms on pro-social behavior. *J. Econ. Psychol.* **30**, 307–320 (2009).
45. A. P. Fiske, The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychol. Rev.* **99**, 689–723 (1992).
46. E. U. Weber, Climate change demands behavioral change: What are the challenges? *Soc. Res. Int. Q.* **82**, 561–580 (2015).
47. Deutsche Gesellschaft für Psychologie (DGPs), *Ethisches Handeln in der psychologischen Forschung Empfehlungen der Deutschen Gesellschaft für Psychologie für Forschende und Ethikkommissionen* (Hogrefe, 2018).
48. A.-C. Posten, P. Ugurlar, S. Kube, J. Lammers, Vertical trust maintains cooperation. OSF. <https://doi.org/10.17605/OSF.IO/DUVRC>. Deposited 30 November 2024.
49. J. Andreoni, L. K. Gee, Gun for hire: Delegated enforcement and peer punishment in public goods provision. *J. Public Econ.* **96**, 1036–1046 (2012).