Examining consequentialist punishment motives in one-shot social dilemmas

Ethics Statement:

Our study is in line with the ethical guidelines of the German Association for Psychology ('Deutsche Gesellschaft für Psychologie') and will be sent out for ethical assessment to the local ethics committee of the Psychology Department of the University of Cologne.

Abstract

Research question

We contrast the intuitive retributivist theory with an alternative, consequentialist theory of punishment by investigating whether consequentialist motives underlie punishment decisions in single round social dilemmas in which there is no prospect of reciprocity. So far, punitive behavior in these one-shot interactions has been taken as an indicator for the presence of retributive motives, since the structure of a one-shot game determines that interaction partners will not interact again. To the best of our knowledge, however, this interpretation has never been tested empirically. We do so by experimentally manipulating whether punishers receive information on the effectiveness of punishment (i.e., information that indicates potential regret on the part of the transgressor) and by investigating how this affects people's willingness to punish. Additionally, we take person-situation interactions into account and examine whether punishment is stronger for prosocial individuals (i.e., persons high in Honesty-Humility and Social Value Orientation) when they receive consequentialist information.

Study methods

The experiment is going to be conducted in accordance with economic standards. That is, interactions are real, without any deception. Decisions (e.g., contributions to the public good, punishment investments) are incentivized, meaning participants are paid according to the decisions they made.

Introduction

The question whether punishment is conducted out of retributive or consequentialist motives represents an ongoing debate and has also been investigated in the context of social dilemmas and economic games. Social dilemmas are defined as situations in which the individual interest conflicts with the collective interest, as for instance when it comes to providing and maintaining public goods (Dawes, 1980; Van Lange, Joireman, Parks, & Van Dijk, 2013). As insufficient cooperation behavior in social dilemmas is often punished (e.g., Fehr & Gächter, 2002), such economic games (e.g., a public goods game, see Ledyard, 1995) can be used to investigate how retribution (i.e., an act of revenge to retaliate free-riding) and consequentialist motives (i.e., to deter from further free-riding) contribute to punishment. For instance, punishment in single-round (i.e., one-shot) interactions has traditionally been assumed to mostly reflect retributive motives as there are no future interactions to educate the offender for (Crockett, Özdemir, & Fehr, 2014; Mischkowski, Glöckner, & Lewisch, 2018). Here, we aim to critically test this notion by studying whether the availability of consequentialist information influences people's punitive behavior in a one-shot economic game.

The current study conceptually replicates research on how consequentialist aspects, as opposed to mere retributivist aspects, affect punishment-related outcomes. Specifically, we relate our study to Funk, McGeer, & Gollwitzer (2014), Study 1, where participants report being more satisfied if punishment is followed by feedback from the transgressor compared to when punishment lacks any kind of feedback. Instead of assessing hedonic reactions after punishment, we look at the precedent punishment behavior as dependent variable. That is, we test if the availability of receiving feedback from the transgressor already manifests itself in actual punishment decisions. Previous research using one-shot economic games suggests that punishers adjust their punishment behavior depending on the consequences that punishment may have. In research on "hidden" versus "open" punishment, Crockett and colleagues (2014) found that participants are more likely to punish and punish more harshly, if their punished partners will find out that their final pay-off was the result of a punitive reduction (i.e., if punishment was "open"). Put differently, when the punished partners would only find out about their final pay-off without any information about how it came about (i.e., if punishment was "hidden"), participants punished less. Crockett and colleagues interpret this difference as the effect of deterrent (i.e., consequentialist) motives in the "open" punishment condition, or a concern about communicating

norms that accompanies people's retributive motives (as identified in the "hidden" punishment condition).

Also using one-shot interactions, Molnar, Chaudhry, and Loewenstein (2020) found that participants are willing to choose a punishment for their partner that is less severe if, along with the punishment, their partner will read why their bonus has been reduced (e.g., "because you were unfair to your partner in the previous task"). Thus, punishers want transgressors to know that they are being punished and why (for similar findings from social psychology, see Gollwitzer & Denzler, 2009; Gollwitzer, Meder, & Schmitt, 2011). People also punish less, for instance, if they can communicate their emotions (Xiao & Houser, 2005), and punishers expect harmless punishment to be as effective as harmful punishment if it is communicative (Sarin, Ho, Martin, & Cushman, 2020; see also Cushman, Sarin, & Ho, in press).

The present study looks at a related yet different facet of consequentialist punishment and manipulates a factor that has not been studied yet in the context of punishment decisions. Instead of manipulating punishers own possibility to express themselves, we vary whether punishers are able to get feedback about the consequences that their punishment has for the transgressor. In previous research, participants' choice to punish in a way that includes an explanation was slightly related to how much participants agreed that such a punishment – even if less severe – would make their partner behave better in the future (Molnar et al., 2020). Punishers care about knowing whether punishment has affected a transgressor or not, be it in a positive or negative way. In studies related to punishers' hedonic reactions, people expect punishment to be more satisfying if they imagine to receive feedback from the transgressor compared to when they imagine to receive no feedback from the transgressor after punishment; and this difference in satisfaction can also be found if participants actually receive feedback or not (Funk et al., 2014, Study 1). Plus, the content of the feedback matters, such that information about a positive transgressor change after punishment makes punishment more satisfying than feedback about no change (see Funk, Gollwitzer, & McGeer, 2014, Study 2). Yet whether the availability of such consequentialist information affects people's punitive behavior to begin with (and therefore potentially motivates it) has not been studied so far.

In the proposed study we will experimentally vary whether participants are able to find out if their punishment has had an effect on the transgressor. Keeping people's retributive and expressive options constant between conditions, we examine how the availability of this kind of

consequentialist information (i.e., the prospect that punishers will get to know which effect their punishment had on the transgressor) affects people's punitive reactions. If punishment in one-shot interactions was purely retributive, the availability of information on the effect of one's punishment should not influence the degree of punishment. However, assuming consequentialist motives do contribute to punishment decisions in one-shot interactions, we expect punishment behavior to increase when the corresponding consequentialist motives are addressed.

To account for individual differences, we assess dispositional prosociality that might moderate the effect of information availability on people's punishment decision. Specifically, we hypothesize increased consequentialist punishment motives for dispositional prosocials as these individuals should be more inclined to punish to establish prosocial norms that prevent them from (further) being exploited in the future. As in this study consequentialist motives are only addressed when the corresponding information is provided, we expect increased punishment behavior in the respective condition for dispositional prosocials as compared to when no consequentialist information is provided. As proself individuals per definition only maximize their own welfare, they should invest few monetary resources to punish, independent of whether consequentialist information is provided or not.

To assess dispositional prosociality, we rely on Social Value Orientation (SVO) as a measure of social preferences, next to Honesty-Humility as the related basic trait dimension of the HEXACO personality inventory (Ashton & Lee, 2007; Lee & Ashton, 2004). Specifically, SVO (Van Lange, 1999; Murphy, Ackermann, & Handgraaf, 2011) consists of an individual difference measure "defined in terms of the weights people assign to their own and others' outcomes in situations of interdependence" (Balliet, Parks, & Joireman, 2009, p. 533). It is operationalized as a series of Dictator Games in which individuals allocate monetary resources between themselves and an anonymous other person. Even though SVO has been shown to be highly predictive of cooperation behavior in social dilemmas (for a recent meta-analysis, see Pletzer, Balliet, Joireman, Kuhlman, Voelpel, & Van Lange, 2018), there is heterogeneous evidence on whether prosocials (i.e., individuals who consider the other person's outcome when allocating resources) punish more harshly than proself (i.e., selfish) individuals. While some studies do not find a difference between prosocials and proselfs (e.g., Böckler, Tusche, & Singer, 2016; Mischkowski et al., 2018; Yamagishi et al., 2012), others do find such a difference – both, in the expected direction of increased punishment for prosocials (e.g., Bieleke, Gollwitzer, Oettingen, & Fischbacher, 2016;

Haruno, Kimura, & Frith, 2014) as well as in the reverse direction of decreased punishment for prosocials (Karagonlar & Kuhlman, 2013). We attempt to explain this heterogeneous evidence by identifying a potential boundary condition. We investigate whether prosocials punish more strongly especially when consequentialist motives are addressed. Specifically, using the materials and measures of Mischkowski et al. (2018) our methods allow for a direct replication of the relation between SVO and punishment investments in the no feedback condition. We expect to replicate their identified null effect of SVO on punishment investments in the no feedback (i.e., baseline) condition. However, when consequentialist feedback will be provided (i.e., in our experimental condition), we expect prosocials to punish more strongly than proselfs – thereby explaining heterogeneous evidence on the relation between SVO and punishment.

To validate the person-situation interaction with a related, yet broader individual difference measure, we investigate whether a similar interaction pattern holds for individuals high in Honesty-Humility. Honesty-Humility "represents the tendency to be fair and genuine in dealing with others, in the sense of cooperating with others even when one might exploit them without suffering retaliation" (Ashton & Lee, 2007, p. 156). It thus represents a form of active (vs. reactive) prosociality and has been shown to be related to SVO (Hilbig, Glöckner, & Zettler, 2014). As individuals high in SVO and Honesty-Humility cooperate in the first place (e.g., Balliet et al., 2009; Hilbig et al., 2013), they inherently face a risk of being exploited. As this rationalizes consequentialist punishment to establish prosocial norms, we hypothesize that prosocials as measured by their SVO and Honesty-Humility punish more severely when consequentialist motives are addressed as compared to when they are not. In turn, we expect generally low punishment investments for individuals low in SVO and Honesty-Humility.[1]

Hypothesis, aims and objectives

In the current study, we examine whether participants invest more resources to punish (i.e., punish more harshly) if they are able to find out its effects on the transgressor.

[1] It is important to note that Agreeableness, rather than Honesty-Humility, has been shown to be negatively related to punishment behavior (Hilbig, Thielmann, Klein, & Henninger, 2016; Thielmann, Spadaro, & Balliet, 2020). However, we focus less on the general relation between prosocial traits and punishment, but on potentially increased consequentialist punishment motives for (active) prosocials (i.e., individuals high in SVO and Honesty-Humility) who face a risk of being exploited in comparison to reactive prosocials (i.e., individuals high in Agreeableness). We report bivariate correlations between all HEXACO dimensions and punishment investments as well as the interaction effect of each HEXACO dimension with the feedback condition in an online appendix

H1: If punishment is not purely retributive but also driven by consequentialist motives, participants who *receive information on the consequences of punishment on the transgressor* invest more resources to punish than participants who know they will not receive this information.

Additionally, taking potential person-situation interactions into account offers an even further differentiated perspective on punishment behavior and its underlying motives. We hypothesize that the effect of information availability on punishment behavior is more pronounced for prosocial individuals (see Figure 1).

H2a: There is an interaction between SVO and whether participants *receive information on the consequences of punishment* on people's punishment behavior, reflecting that prosocials increase their punishment investments when consequentialist motives are addressed. That is, we expect prosocials to show stronger punishment in the condition with available information in comparison to the condition without given information. We expect a lower increase – if any – in the punishment behavior of proselfs when consequentialist information is provided as compared to the control condition without this information.

H2b: We expect a similar interaction pattern as outlined in H2a for individuals high in Honesty-Humility.

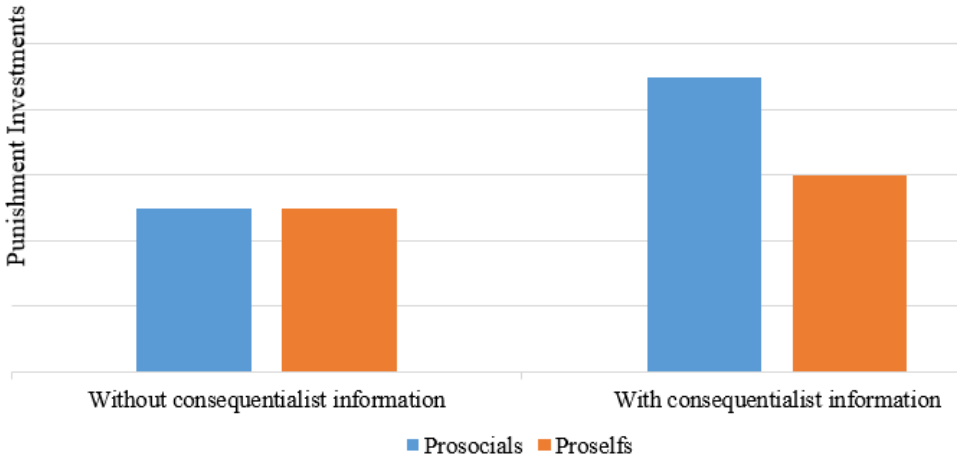

Figure1. Illustration of the hypothesized interaction pattern (H2). Punishment investments as a function of the availability of consequentialist information and dispositional prosociality.

Materials and Methods

All materials, codes and anonymized data will be made public after data collection on the open science framework (https://osf.io/). Instructions are attached to this submission (see Appendix I).

Data collection

Data will be collected via the Decision Lab Cologne (DLC), the data base of the Social Psychology research group at the University of Cologne (UoC). The data base mainly consists of students from the UoC above 18 years who registered at the data base to receive invitations to take part in psychological studies. The data base is built in accordance with recent European data protection regulations ('DSGVO') and approved by the ethics committee of the Department of Psychology. Specifically, participants were informed and agreed in the personality base assessment that their personality data is going to be linked to their data of subsequent studies.

For the current study, lab regulations due to the current corona virus pandemic render data collection in the lab impossible. Therefore, participants will be invited via email to take part in the experiment online. The invitation entails information about the estimated duration, average payment and broad information about the study's content. Those participants who subscribed to a virtual session and completed the study are paid in accordance with their and their group members' decision-making. The study itself will be run via oTree (Chen, Schonger, & Wickens, 2016).

Sample size, power and precision

For each hypothesis, we conducted an a priori power analysis in G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) based on an alpha level of .05 and a power of .90. Relying on the identified medium effect size (d = 0.52) of Funk et al. (2014, Study 1, Hypothesis b) regarding the group comparison between imagined feedback and indicated satisfaction after punishing, we conservatively assume a small to medium-sized difference (f^2 = .10) for the influence of feedback availability on punishment behavior (H1) and an even smaller effect size (i.e., f^2 = .08) for the interaction effect between feedback and dispositional prosociality (H2). The regression model to test the first hypothesis includes two predictors: the independent variable information availability (yes / no) as well as an important covariate: the averaged contribution behavior of the other group members relative to one's own contribution to assess the effect of

information availability independent of the degree of experienced exploitation. The power analysis reveals a necessary sample size of $N = 130$.

The second hypothesis investigates whether the availability of information on the consequences of punishment increases the punishment investments for prosocials (i.e., for individuals high in SVO (H2a) and high in Honesty-Humility (H2b)).[2] As outlined, we assume a smaller effect size than the main effect (i.e., $f^2 = .08$). The corresponding regression models includes four predictors (information availability, the individual difference moderator (SVO / Honesty-Humility), the interaction term of SVO / Honesty-Humility and information availability as well as the averaged contribution behavior of the other group members). This power analysis reveals a necessary sample size of $N = 198$.[3]

To assure a high power for the interaction effect, we will collect data of $N = 200$ participants. We will stop data collection as soon as the number of participants is reached, but no later than two weeks before the submission deadline of the stage 2-protocol to be able to analyze the data and complete the manuscript.

Participant characteristics

Participants registered at the Decision Lab Cologne – the subject pool for this study – are above 18 years and consist mainly of students from University of Cologne from all academic disciplines. We therefore expect a typical student sample with an average age in the mid-twenties and roughly balanced gender distribution.

Procedure

In general, when registering as a participant at the Decision Lab Cologne, participants first take part in a base assessment, containing a broad range of personality questionnaires (e.g., basic traits, social preferences, cognitive reflection capacity, etc.). As data is pseudonymized via a Decision Lab ID, participants have consented that data can be linked to subsequent studies. Thus, the relevant personality data (i.e., SVO and the HEXACO personality inventory containing Honesty-Humility) is already collected, allowing us to solely run the Public Goods Game for this study.

[2] As Honesty-Humility and SVO are correlated (see e.g., Hilbig et al., 2014), we will conduct two regressions and separately include SVO and Honesty-Humility (keeping the model otherwise constant) to independently test our hypotheses H2a and H2b.

[3] Note that this requires to collect at least $N = 200$ participants as the PGG consists of groups with four players.

Focusing on the detailed procedure of the data collection, participants first provide informed consent and receive the instructions of a public goods game in groups of four players, including corresponding comprehension questions (for details, see section on exclusion criteria, plus see instructions in Appendix I). After having passed this comprehension check, participants are randomly assigned to one of the two conditions (with or without information of the effects of punishment on the transgressor). Participants decide how much to contribute to the public good and decide whether and, if applicable, how much money to subtract from at least one of their group members (i.e., second-party punishment). Participants in the feedback conditions answer two questions about whether they would change their contribution level in a subsequent interaction and whether they regret their previous behavior, serving as consequentialist information. When having completed the public goods game, participants are informed about their final outcome, thanked and debriefed.[4]

Conditions and design

We test our hypotheses in a one-factorial between-subject design, manipulating whether information about the consequences of one's punishment on the transgressor is provided (or not).

Manipulated and measured variables

As outlined in the condition section, our manipulated variables focus on the *availability of information on the consequences of punishment* (IV: feedback condition).

To assess whether the impact of consequentialist motives in one-shot interactions differs among prosocials and proselfs, we measure SVO (H2a) via the 15 items SVO Slider Measure (Murphy et al., 2011) and Honesty-Humility (H2b) as part of the HEXACO personality inventory (100 items version; Lee & Ashton, 2018). As dependent variable, we assess the punishment magnitude (i.e., punishment investments).

As further measured variables, but not used for confirmatory analyses, we assess cooperation behavior as individual contributions to the public goods game. In addition, we assess in the feedback condition the consequentialist information for each group member in the public goods game. Specifically, we ask participants whether they would behave differently in a similar future interaction and whether they regret their contribution behavior.

[4] Study participation at the Decision Lab Cologne is paid via bank transfer a few days after having participated.

Analysis Plan

Preprocessing

Exclusion criteria. To assure sufficient game understanding, control questions to the public goods game are asked at the very beginning of the study and need to be answered correctly in order to continue with the study. Participants read the description of the public goods game and are asked which contribution a) maximizes their own outcome (correct response: none at all), b) maximizes the group's outcome (correct response: contributing everything) and c) how much they have to invest to subtract 1€ from a group member (correct response: 25 Cents, see Appendix I). Up to three incorrect answers are allowed; after the third incorrect response instructions are displayed again (the relevant parts being marked in bold). In case some participants are still not able to answer the control questions correctly, they are allowed to take part, but their data will be excluded afterwards.[5]

Secondly, we ask participants as a manipulation check at the very end of the study about the characteristics of the condition they were assigned to (i.e., which information participants could receive about their group members, see Appendix I). We will not exclude participants who answer incorrectly, but will separately report the results of our confirmatory analyses for both subgroups (participants with a successful and unsuccessful manipulation check). In case results do not differ when including the incorrect answers of the manipulation check, the full sample is reported.

Transformations. Even when punishing to a similar extent (i.e., investing the same amount on absolute terms), the punishment investment differs in its severity for each individual dependent on the outcome of the first stage of the public goods game (i.e., after each of the four group members made their contributions). To account for this relative difference, punishment investments are transformed as percental investment to the earnings from the public good. Otherwise, no transformations are made. Note that punishment investments as a continuous variable also contain the binary decision, whether or not to punish at all, as the no-punishment decisions are coded as zero (Cents) investment.

Tests

[5] As a robustness check, we will analyze whether our results also hold when including individuals with a lack of game understanding.

To test the first hypothesis (H1), that participants who *receive information on the consequences of punishment on the transgressor* punish more in that they invest more resources to punish than participants who know they will not receive this information, punishment investments will be regressed on the feedback condition. Importantly, we will control for the averaged group members' contribution behavior relative to the own contributions to the public good (i.e., the difference between own and averaged group members' contributions) as this reflects the experienced wrongdoing that most likely influences the severity of punishment. By keeping its influence constant, we can independently assess the effect of information availability.

To test the interaction effect (H2), that prosocials punish more severely when consequential motives are addressed as compared to when they are not, we will conduct two analyses. First, we will regress the punishment magnitude on the experimental manipulation of information availability, SVO and their interaction term (H2a), controlling for the averaged group members' contribution behavior (see above). SVO will be centered on the sample mean. We see our hypothesis as confirmed if there is a significant interaction, reflecting a higher punishment magnitude for prosocials in comparison to proselfs when consequentialist information is given. Second, we will conduct the same regression model with Honesty-Humility (H2b) instead of SVO to analyze whether the interaction potentially persists when operationalizing dispositional prosociality with a broader basic trait. We see this hypothesis as confirmed if there is a similar, statistically significant interaction pattern as described for SVO.

References

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11*(2), 150-166. https://doi.org/10.1177/1088868306294907

Balliet, D., Parks, C., & Joireman, J. (2009). Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Processes and Intergroup Relations*, *12*(4), 533-547. https://doi.org/10.1177/1368430209105040

Bieleke, M., Gollwitzer, P. M., Oettingen, G., & Fischbacher, U. (2016). Social value orientation moderates the effects of intuition versus reflection on responses to unfair ultimatum offers. *Journal of Behavioral Decision Making*, *30*(2), 569–581 https://doi.org/10.1002/bdm.1975

Böckler, A., Tusche, A., & Singer, T. (2016). The structure of human prosociality: Differentiating altruistically motivated, norm motivated, strategically motivated, and self-reported prosocial behavior. *Social Psychological and Personality Science*, *7*(6), 530-541. https://doi.org/10.1177/1948550616639650

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree - An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance, 9*, 88-97. https://doi.org/10.1016/j.jbef.2015.12.001

Crockett, M. J., Özdemir, Y., & Fehr, E. (2014). The value of vengeance and the demand for deterrence. *Journal of Experimental Psychology: General, 143,* 2279-2286. https://doi.org/10.1037/xge0000018

Cushman, F.A., Sarin, A. & Ho, M. (in press). Punishment as communication. In: J. Doris & M. Vargas (Eds.). *Oxford Handbook of Moral Psychology*. Oxford University Press.

Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology*, *31*(1), 169-193. https://doi.org/10.1146/annurev.ps.31.020180.001125

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*, 137-140. https://doi.org/10.1038/415137a

Feinberg, J. (1965). The expressive function of punishment. *The Monist, 49*, 397-423.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Funk, F., McGeer, V., & Gollwitzer, M. (2014). Get the message: Punishment is satisfying if the transgressor responds to its communicative intent. *Personality and Social Psychology Bulletin, 40*(8), 986-997. https://doi.org/10.1177/0146167214533130

Gollwitzer, M., Denzler, M. (2009). What makes revenge sweet: Seeing the offender suffer or delivering a message? *Journal of Experimental Social Psychology, 45*(4), 840-844. https://doi.org/10.1016/j.jesp.2009.03.001

Gollwitzer, M., Meder, M., Schmitt, M. (2011). What gives victims satisfaction when they seek revenge? *European Journal of Social Psychology, 41*(3), 64-374. https://doi.org/10.1002/ejsp.782

Haruno, M., Kimura, M., & Frith, C. D. (2014). Activity in the nucleus accumbens and amygdala underlies individual differences in prosocial and individualistic economic choices. *Journal of Cognitive Neuroscience*, *26*(8), 1861-1870. https://doi.org/10.1162/jocn_a_00589

Hilbig, B. E., Glöckner, A., & Zettler, I. (2014). Personality and pro-social behavior: Linking basic traits and social value orientations. *Journal of Personality and Social Psychology, 107*(3), 529–539. https://doi.org/10.1037/a0036074

Hilbig, B. E., Thielmann, I., Klein, S. A., & Henninger, F. (2016). The two faces of cooperation: On the unique role of HEXACO Agreeableness for forgiveness versus retaliation. *Journal of Research in Personality*, 64, 69-78. https://doi.org/10.1016/j.jrp.2016.08.004

Kahneman, D., Ritov, I., Jacowitz, K. E., & Grant, P. (1993). Stated willingness to pay for public goods: A psychological perspective. *Psychological Science*, *4*(5), 310-315. https://doi.org/10.1111/j.1467-9280.1993.tb00570.x

Karagonlar, G., & Kuhlman, D. M. (2013). The role of social value orientation in response to an unfair offer in the ultimatum game. *Organizational Behavior and Human Decision Processes*, *120*(2), 228-239. https://doi.org/10.1016/j.obhdp.2012.07.006

Ledyard, J. O. (1995). Public goods: A survey of experimental research. In J. H. Kagel & A. E. Roth (Eds.), *The handbook of experimental economics* (pp. 111-194). Princeton, NJ: Princeton University Press.

Lee, K., & Ashton, M. C. (2018). Psychometric Properties of the HEXACO-100. *Assessment, 25*(5), 543-556. https://doi.org/10.1177/1073191116659134

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research, 39*(2), 329-358. https://doi.org/10.1207/s15327906mbr3902_8

Liebrand, W. B., Jansen, R. W., Rijken, V. M., & Suhre, C. J. (1986). Might over morality: Social values and the perception of other players in experimental games. *Journal of Experimental Social Psychology*, *22*(3), 203-215. https://doi.org/10.1016/0022-1031(86)90024-7

Mischkowski, D., Glöckner, A., & Lewisch, P. (2018). From spontaneous cooperation to spontaneous punishment – Distinguishing the underlying motives driving spontaneous behavior in first and second order public good games. *Organizational Behavior and Human Decision Processes, 149*, 59-72. https://doi.org/10.1016/j.obhdp.2018.07.001

Molnar, A., Chaudhry, S., & Loewenstein, G. F. (2020, January 24). 'It's Not About the Money. It's About Sending a Message!': Unpacking the Components of Revenge. http://dx.doi.org/10.2139/ssrn.3524910

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. J. J. (2011). Measuring social value orientation. *Judgment and Decision Making*, *6*(8), 771–781. Retrieved from http://journal.sjdm.org/11/m25/m25.pdf

Pletzer, J. L., Balliet, D., Joireman, J., Kuhlman, D. M., Voelpel, S. C., & Van Lange, P. A. (2018). Social value orientation, expectations, and cooperation in social dilemmas: A meta-analysis. *European Journal of Personality*, *32*(1), 62-83. https://doi.org/doi.org/10.1002/per.2139

Sarin, A., Ho, M. K., Martin, J., & Cushman, F. A. (2020, March 26). Punishment is Organized around Principles of Communicative Inference. https://doi.org/10.31234/osf.io/2cyf7

Thielmann, I., Spadaro, G., & Balliet, D. (2020). Personality and prosocial behavior: A theoretical framework and meta-analysis. *Psychological Bulletin*, 146(1), 30-90. https://doi.org/10.1037/bul0000217

Van Lange, P. A. M. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology*, *77*(2), 337-349. https://doi.org/10.1037/0022-3514.77.2.337

Van Lange, P. A. M., Joireman, J. A., Parks, C. D., & van Dijk, E. (2013). The psychology of social dilemmas: A review. *Organizational Behavior and Human Decision Processes, 120,* 125-141. https://doi.org/10.1016/j.obhdp.2012.11.003

Xiao, E., & Houser, D. (2005). Emotion expression in human punishment behavior. *PNAS*, *102*, 7398-7401. https://doi.org/10.1073/pnas.0502399102

Yamagishi, T., Horita, Y., Mifune, N., Hashimoto, H., Li, Y., Shinada, M., . . . Simunovic, D. (2012). Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity. *Proceedings of the National Academy of Sciences*, *109*(50), 20364-20368. https://doi.org/10.1073/pnas.1212126109