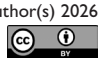
*Article*

# Receiving Applause "From the Right": Punishing Freeriding Is Socially Rewarded by Those High in Right-Wing Authoritarianism

Personality and Social Psychology Bulletin
1–10


Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/01461672261445479
journals.sagepub.com/home/pspb


**Richard Rau[1*], Luise Hönig[2,3*], Mitja D. Back[2], and Michael P. Grosz[1]**

## Abstract

Most social groups punish freeriders (i.e., individuals who receive the same benefits from the group as others, despite contributing less to its success). In small groups, individual group members (rather than established authorities) typically implement this punishment spontaneously, and punishers may consequently be awarded social status by their peers. Here, we tested the preregistered hypothesis that this way of acquiring status works best when fellow group members are high, rather than low, in right-wing authoritarianism (RWA). The hypothesis was supported in a laboratory-based behavioral study ($N = 667$) in which small groups interacted in a financially incentivized repeated public goods game involving punishment (i.e., a social dilemma that puts immediate individual benefits at odds with long-term collective interests). Linking the process of status acquisition to peer RWA significantly advances the understanding of social dynamics in small groups.

## Keywords

right-wing authoritarianism, freeriding, punishment, social status, public goods game, social relations model

Received: October 15, 2025; accepted: April 3, 2026

## Introduction

Any group, organization, or society requires its members to act cooperatively to achieve a safe and prosperous environment for everyone. At the same time, individuals naturally have an incentive to freeride, that is, to benefit from cooperative social groups while individually contributing less than others to them (Fischbacher & Gächter, 2010).[1] One way to mitigate freeriding is to establish explicit norms (e.g., laws) and formal authorities (e.g., police, courts) that punish those who violate these norms. However, norm violations and punishments occur not only in society at large but also in smaller social groups, which may include both pre-existing groups (e.g., discussion boards, work teams) and groups that are formed ad hoc (players in randomly matched online games, travel groups). In these less formalized settings, freeriding punishment is often carried out spontaneously by other group members rather than by formal authorities, even when this involves individual costs for the punisher. While freeriding punishment quite effectively maintains a cooperative group dynamic that benefits the group as a whole (Fehr & Gächter, 2002; Henrich & Boyd, 2001), the question of whether and when punishers also benefit individually from the act of punishing is more complex (Gordon & Lea, 2016; Raihani & Bshary, 2015; Redhead et al., 2021). On the one hand, their behavior may be considered morally virtuous, and virtues have been theorized to lead to social status (Bai, 2017; Grosz et al., 2025b). In line with this, some studies have suggested that punishment increases the punisher's social status (Barclay, 2006; Jordan et al., 2016; Nelissen, 2008). However, other studies have failed to support this finding (Grosz et al., 2025b; Horita, 2010; Ozono & Watabe, 2012). The reason for this inconsistency might be that punishment also has an aversive character, as it involves aggression or degradation and may be seen as morally questionable (Eriksson et al., 2016). Here, we propose that individual differences in *right-wing authoritarianism* (RWA) among those granting status to others after having witnessed them (not)

[1]HMU Health and Medical University, Potsdam, Germany
[2]University of Münster, Germany
[3]University of Vienna, Austria

*These authors shared first authorship.

**Corresponding Author:**
Luise Hönig, University of Münster, Fliednerstraße 21, Münster 48149, Germany
Email: luise.hoenig@univie.ac.at

punishing freeriding may explain when status benefits occur and when they do not.

Rooted in sociological work on the origins of fascism and antisemitism (Adorno et al., 1950), RWA is defined as a conservative and authority-oriented attitude, where people high in RWA have a powerful desire to maintain the values of a society and in-group cohesion (Altemeyer, 1981, 1998; Gerber & Jackson, 2015). On the basis of this desire, individuals high in RWA might be particularly willing to grant social status to those who punish freeriders, that is, to admire them, perceive them as instrumentally valuable, and comply with them (Anderson et al., 2015). If correct, our hypothesis would suggest that inconsistent findings on the link between punishment and status attainment in previous research might be explained by between-study differences in participants' RWA.

## Why Individuals High in RWA Value Punishment

RWA spans the facets of submission to authorities, adherence to the conventions of a society, and authoritarian aggression (Altemeyer, 1981). Common to all three facets is a perception of social threat (Duckitt & Sibley, 2009), which includes both realistic and symbolic threats (Stephan et al., 1999). As a means to reduce these threats and restore collective security, people high in RWA have been found to support measures such as restrictions on civil liberties (Cohrs et al., 2005), a hard-charging police force (Gerber & Jackson, 2017; Saunders et al., 2016), and harsh legal sentences for criminal offenders (Gerber & Jackson, 2013, 2015). Whereas this previous research explains who tends to support drastic measures to prevent norm violations at the societal level, it does not speak directly to the issue of freeriding in smaller social units with less formalized social orders. For example, in collaborative online Q&A communities, implicit norms emerge organically even though no formal authority enforces them (Burnett & Bonnici, 2003; Liao et al., 2020). Do individuals high in RWA also support ad hoc punishment coming from another group member in these scenarios? We hypothesized that they do. On the basis of their desire to restore social order, high RWA individuals should be inclined to emphasize the morally virtuous aspect of punishment and appreciate group members who step up and volunteer to sanction uncooperative behavior. On the flipside, low RWA individuals should be inclined to see punishment as aversive and appreciate it less.

## The Present Research

To our knowledge, studies on the link between freeriding punishment and status attainment have yet to address individual differences in RWA in a target person's fellow group members (i.e., among those who actually assign status to the target person). Likewise, few studies on the link between punishment and RWA have addressed (a) small groups (as opposed to society at large) and (b) freeriding punishment by laypersons (as opposed to authorities such as police officers; see Gerber & Jackson, 2013). This research gap is unfortunate given that many everyday freeriding conflicts pertain to small groups, which lack explicit norms and authorities. The present work is aimed at understanding the role of group members' RWA in such small group settings.

Participants in the present study completed a self-report measure of RWA online and several days later showed up to participate in a laboratory session involving groups of seven to nine unacquainted individuals. In this session, they interacted via computers in an iterative public goods game with punishment that involved real monetary incentives. For some group members, this game offered the opportunity to freeride, and for others, it offered the opportunity to punish freeriding at their own cost. Social status was assessed via mutual ratings and a ranking. We hypothesized that a participant's RWA would moderate the extent to which their status assignments were contingent upon the extent to which group members had punished freeriders during the public goods game.

> **H1:** The higher an individual's level of RWA, the more positive the association between the social status they allocate to different target individuals and these target individuals' punishment of freeriders.

The moderating effect described in H1 may be pronounced for the authoritarian aggression facet of RWA because this facet describes the extent to which a person supports punishing norm violations (Altemeyer, 1981).

> **H2:** The higher an individual's level of authoritarian aggression (i.e., a facet of RWA), the more positive the association between the social status they allocate to different target individuals and these target individuals' punishment of freeriders.

## Method

We analyzed data from the Group Interaction and Perception of Others (GIPO) project, a larger project designed to address a variety of research questions. We describe only procedures and materials that are directly relevant to the current study. Comprehensive documentation of the entire GIPO project is available at https://osf.io/e3cpn (Grosz et al., 2025a). The hypotheses of the present study, including a detailed analysis plan, were preregistered before data collection was complete. In a few instances, deviations from the registration became necessary during data analysis. The preregistration, a list of deviations (see Supplemental Table S1), additional results (Supplemental Tables S2 and S3), and data and *R*-code for reproducing all reported results can be retrieved from https://osf.io/6enp4. All analyses were performed in *R* (Version 3.6.1; R Core Team, 2019).

### Participants

We recruited German speakers between 18 and 55 years of age via social media postings, flyers, and announcements during university lectures. Interested individuals completed an online survey, which, besides the measure of RWA and other scales, included three attention-check items (e.g., "Please choose the response category 'does not apply at all' here"). Only participants who responded as instructed on at least two of the three items were scheduled for the laboratory experiment. A sample of $N = 667$ participants (84 groups, $M_{age} = 22.47$, $SD_{age} = 4.54$) provided usable data during the laboratory session.[2] The majority of them identified as female (63% female, 36% male, 1% diverse). Participation was compensated either with course credit or with 12€. In addition, participants received financial payoffs from the public goods game (see below) and from other tasks that were part of the larger project but are not relevant to the present study.

### Measures

To assess RWA, the online survey included the Authoritarianism Short-Scale KSA-3, which includes three items for each subscale, rated from 1 (*do not agree at all*) to 5 (*agree completely*; Beierlein et al., 2014). We computed scale scores by averaging the respective items, which produced reliable scores for both the overall scale ($\alpha = .81$) and the authoritarian aggression subscale ($\alpha = .78$).

Punishment and social status were assessed in the laboratory session. Punishment was operationalized as a person's overall amount of money invested in punishing freeriders. Social status was assessed after the fourth round of the modified public goods game as a dyadic variable (i.e., pertaining to a unique combination of perceiver and target). To capture the aspects of admiration and perceived instrumental value of the social status definition (Anderson et al., 2015), participants rated the items "I admire Person X" and "Person X is useful to achieve my goals in the group" on a scale ranging from 1 (*do not agree at all*) to 6 (*agree completely*). The aspect of voluntary compliance was assessed via the election of one group member as someone who would have control over everyone's payoffs in a subsequent extra round of the public goods game (see the "Laboratory Session" section below). Specifically, participants assigned ranks to all group members except for themselves, ranging from 1 (*most preferred in a position of power*) to sixth, seventh, or eighth (*least preferred*), depending on the group size. The consistency of the three aspects of social status was moderate ($\alpha = .66$), and we ran supplementary analyses for each aspect separately.

Our analyses did not consider status assignments from dyads in which the perceiver was a freerider, and the target was a punisher of that particular instance of freeriding, because these ratings and rankings are arguably driven more by personal retaliation rather than the perceiver's tendency to (dis)approve of third-party punishment. We failed to list the exclusion of these cases as an explicit step in data preprocessing in our preregistration (cf. Supplemental Table S1), but we note that this deviation is consistent with the preregistration mentioning the goal of examining status assignments for those with an observer perspective.

### Laboratory Session

In each session, groups interacted with each other in a computer laboratory with nine computers, separated by visual barriers. Sessions lasted approximately 75 minutes. Upon arriving, participants were assigned a random seat and instructed not to communicate with each other throughout the study. They wore earmuffs to further prevent communication, and due to the COVID-19 pandemic, most of them wore face masks. The experiment was executed on the platform *oTree* (version 2.5.3; Chen et al., 2016). Instructions and visualizations similar to Figures 1 and 2 guided the participants through the entire session. In addition, at the beginning of the lab session, participants received general instructions orally and a printed handout explaining the rules of the first four rounds of the experiment.

The experiment consisted of four rounds of a modified version of the repeated public goods game with punishment (Fehr & Gächter, 2000) and an extra round at the end. For each group size (7, 8, or 9), there were slightly different versions of the experiment. We describe the eight-person version here, but information about the seven- or nine-person versions can be found elsewhere (see https://osf.io/e3cpn). First, every group member was randomly assigned a number from 1 to 8, which they kept throughout the entire session. Then, they completed a training phase, which included an explanation of the games' rules and two trial rounds with preprogrammed fictional group members. After the training, the group members played four rounds of the game in which they interacted with each other. Every round consisted of two stages. In the stage 1, individuals 1 to 4 received 2€ each. They were asked to decide how much of the 2€ they were willing to contribute to the so-called group project and how much they wanted to keep for themselves. For each euro that was contributed to the group project, *all* group members received 0.30€ (see Figure 1). Thus, contributing to the group project was costly for the contributor but beneficial for the other group members. In the second stage of each of the four rounds, individuals 5 to 8 had the opportunity to punish individuals 1 to 4. Specifically, they could decide whether and, if so, how much they wanted to punish each of their fellow group members by investing a specific amount of money, of which the tripled amount would be deducted from this individual's payoff. They could use as much money as they liked from what they had received during the first stage of the current round. The contributions made by individuals 1 to 4 during stage 1 were presented to them as they decided who to

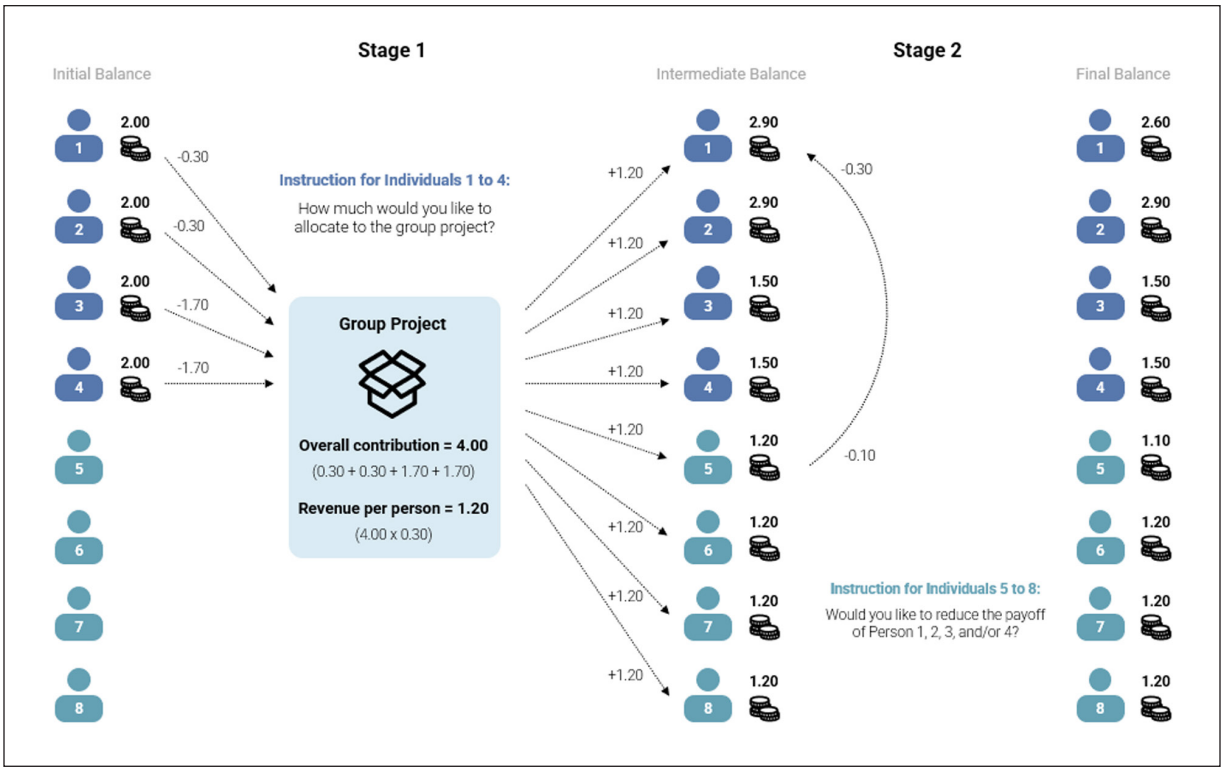



**Figure 1.** Visualization of one regular round of the modified public goods game with the opportunity to punish.
*Note.* Individuals 1 to 4 are contributors, and individuals 5 to 8 are punishers. In stage 1, each contributor gets 2€ and can contribute to the group project (e.g., individual 1 contributes 0.30€). Every participant then receives the same revenue, in this case 1.20€, from the project. During the second stage, individual 5 decides to spend 0.10€ to reduce Individual 1's payoff by 0.30€. The final payoffs can be seen in the right-most part of the figure (e.g., individual 1's final payoff is 2.60€).

punish and how much. After individuals 5 to 8 had made their punishment decisions, all contributions, punishments, and the resulting payoffs for every group member were displayed on the participants' screens.

After rounds 1 and 4, participants provided round-robin ratings. That is, they rated every other group member on admiration and perceived instrumental value. After round 4, they also elected one individual who would take extra responsibility for the group in a final game round by ranking all group members (except for themselves) in terms of how much they preferred to have this individual in this position of power. They were informed that the group member with the lowest overall rank would be tasked to freely reallocate up to 5€ to the group members (including themselves) in a final round of the game. While deciding on their ranking, participants saw a summary of each group member's decisions during the previous four rounds.

In stage 1 of the extra round, *all* eight group members received 2€ and had to decide how much of their money to contribute to the group project. In stage 2, the elected group member could freely reallocate up to 5€ by subtracting specific amounts from some group members' payoffs and adding the same overall amount to others' payoffs (including their own; Figure 2). Notably, data on contribution and reallocation behavior in the extra round are not considered in the present analysis. The sole purpose of the extra round was to create a social situation in which groups put single members in a position of actual power. Once this reallocation was finished, participants were given their earnings, including the payoffs from all five rounds of the game, in an envelope and discharged by the experimenter.

## Statistical Analyses

Our hypothesis concerns a dyad-level dependent variable (i.e., how much social status a perceiver assigns to a target) and proposes an interactive effect of person-level predictors (i.e., between the perceiver's RWA and the target's punishing behavior). However, because we used a round-robin design in which every participant rated and was rated by each fellow group member, our data did not readily lend themselves to standard multilevel modeling. Instead, the dependencies of round-robin data had to be handled by a social relations model (SRMs; Back & Kenny, 2010; Kenny, 1994), which decomposes each dyadic observation (e.g., A admires B) into a *perceiver effect* (how much A admires the typical target), a *target effect* (how much B is admired by the typical perceiver), and a *relationship effect* (how much A specifically admires B after considering A's perceiver effect and B's target effect). Therefore, we removed any person-level

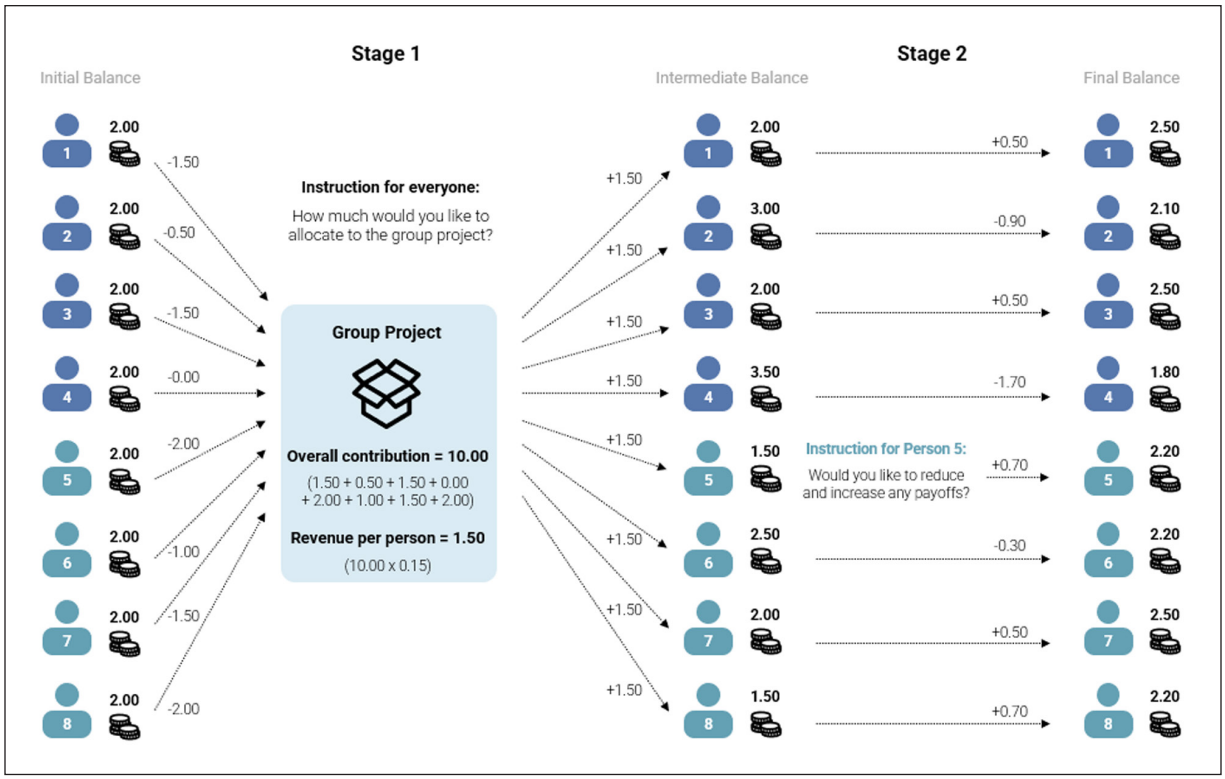



**Figure 2.** Visualization of the extra round.

*Note.* The extra round differed in two respects from the regular rounds. In stage 1, *every* individual could allocate money to the group project. In stage 2, an elected group member could freely reallocate payoffs across all group members. The only constraint was an even balance of the overall payoffs. In the example, individual 4 freerides at a maximum (no contribution), and individual 5 and individual 8 cooperate at a maximum (they contribute the entire 2.00€). In turn, individual 5 (who was elected by the group) reduces individual 4's payoff the most (i.e., by 1.70€) and increases their own and individual 8's payoffs the most (i.e., by 2.00€ each).

dependencies from the data in a preparatory SRM and then used multilevel modeling to analyze data in which only the dyad-level dependencies were left (see Back & Kenny, 2010; Dufner et al., 2024). Specifically, the first step involved running a univariate SRM via *TripleR* (Schönbrodt et al., 2012) for each of the three social status indicators and averaging the standardized relationship effect estimates from these models into a single dyad-level variable. Supplemental Table S2 presents the results for the SRMs.

As anticipated in our preregistration, the resulting data were characterized by small but noticeable dyadic dependencies (intra-class correlation [ICC] = .06), which is why, for our main analysis, we used the *lme4* package (Bates et al., 2015) to fit multilevel regression models with random intercepts for dyads at level 2. The models were estimated via restricted maximum likelihood with a Satterthwaite degrees of freedom approximation for the *t*-tests to control the type I error rate as implemented in the *lmerTest* package (Kuznetsova et al., 2017). The predictors were the target's overall amount of money invested to punish others and the perceiver's grand-mean-centered RWA score. To facilitate interpretation, we report standardized fixed-effect estimates obtained via post-estimation standardization of the multilevel model parameters via *effectsize* package (Ben-Shachar et al., 2020), which can be interpreted as standardized regression coefficients. Due to the multilevel structure and group-mean-centered predictors, standardization may slightly alter estimates; for effects close to zero, this can result in minor changes in magnitude and sign that are not substantively meaningful. Unstandardized effect sizes for the hypotheses and robustness checks are reported in Supplemental Material S3. The regression equation at level 1 was as follows:

$$\begin{aligned}\text{status\_assignment}_{\text{perceiver/target}} &= \text{punishment}_{\text{target}} \\ &+ \text{RWA}_{\text{perceiver}} \\ &+ \text{punishment}_{\text{target}} \\ &* \text{RWA}_{\text{perceiver}} + \varepsilon\end{aligned}$$

The product term models a cross-level interaction to test whether between-perceiver differences in RWA moderate the association between a target's amount of punishment and the social status the target receives from a perceiver. A positive interaction would indicate that RWA is positively associated with the willingness to assign more status to targets who punished uncooperative behavior in the public goods game. To

test H2, we used the authoritarian aggression score instead of the RWA score. Otherwise, the regression equation for H2 was analogous to H1.

In addition, we conducted six robustness checks to get a more nuanced understanding of the stability of our findings. In the first robustness check, we addressed that averaging relationship-effect scores from separate univariate SRMs confounds the composite with indicator-specific measurement error. To mitigate this, we estimated a multiple-indicator multilevel model in which all three status indicators (i.e., admiration, perceived instrumental value, and compliance) were included simultaneously. In this specification, an indicator-specific intercept was included to account for baseline differences between the measures, while the random intercept for dyads captures the shared variance across indicators, which effectively separates true-score variance from indicator-specific error. This approach reduces attenuation of the effects due to measurement error and increases statistical power relative to the averaged-composite approach used in the preregistered main analysis. In this specification, each status indicator contributed an observation of status for a given perceiver–target dyad, and status indicator-specific intercepts were included to account for differences between the status indicators while estimating effects across status indicators. To examine the variance structure of this specification, we first estimated a random-intercept-only model with dyads as the grouping factor. The variance components indicated substantial variation at the dyad level (ICC = .20, $SD$ = 0.44), supporting the use of a multilevel modeling approach.

All subsequent robustness checks were estimated using this multiple-indicator multilevel specification (a). In the second robustness check (b), we included gender as an additional predictor, as preregistered. In the third preregistered robustness check (c), we examined whether the estimated effects differed across the three status indicators by estimating a model that allowed the predictors and their interaction to vary by status indicator (i.e., including interactions between each indicator and the predictors). The fourth robustness check (d) was conducted post hoc and excluded two participants who were older than 50 years, although age 50 had been communicated as the maximum age for participation during recruitment. Fifth, we conducted an additional robustness check that included the perceiver's role (i.e., contributor vs. punisher) as a predictor in model (e). Finally, in the last robustness check (f), we reanalyzed the data without applying one of the data-preparation steps used in the main analyses. In the primary analyses, we excluded status judgments from contributors about punishers who had directly punished them to reduce the possibility that status evaluations were influenced by the experience of being punished (e.g., due to immediate reactions such as resentment or justification). To examine whether this decision affected the results, we conducted an additional analysis in which these status judgments were retained.

## Results

### Descriptive Statistics

Participants' RWA scores ranged from 1.00 to 4.44 ($M$ = 2.07, $SD$ = 0.60), and those in the role of contributors varied greatly in how much they contributed to the group project.[3] Specifically, their overall contributions ranged from 0€ to 8€ (i.e., some participants kept all their money and others gave all their money to the group in all four rounds; $M$ = 6.31€, $SD$ = 1.79€). Freeriding was shown by 71% of the contributors to some extent (i.e., the individuals gave less than 2€ per round). The average contributions observed per round increased slightly over time ($M_{\text{Round1}}$ = 1.39€, $M_{\text{Round2}}$ = 1.61€, $M_{\text{Round3}}$ = 1.66€, $M_{\text{Round4}}$ = 1.67€), hinting that the possibility of punishment might have forestalled a decay in contributions, a phenomenon commonly observed in repeated public goods games with and without punishment (e.g., Fehr & Gächter, 2000). Participants in the role of punishers also varied greatly in how much they spent on punishing others. Their overall spending ranged from 0€ to 5.99€ ($M$ = 0.42€, $SD$ = 0.67€), and 70% of them showed punishment to some extent (i.e., spent more than 0€ overall). Punishment was primarily directed toward freeriders. In particular, of all punishments, 94% were directed at individuals who contributed less than 2€ in the round in which they were punished, and 87% were directed at individuals who contributed less than the group average in that round.

### Main Analysis

The regression model used to test H1 did not yield any significant main effects. That is, targets who invested a lot into punishing others were not generally any more or less likely to acquire high status than were targets who invested smaller amounts, $b$ = 0.04 (standardized), 95% CI [−0.01, 0.08], $p$ = .119. Likewise, perceivers who were higher in RWA were not generally any more or less likely to grant others high status than were perceivers low in RWA, $b$ = 0.01 (standardized) [−0.04, 0.06], $p$ = .517. Crucially, however, there was a significant interaction effect, $b$ = 0.05 (standardized) [0.00, 0.10], $p$ = .038. Thus, as predicted, greater social status was granted to punishers mostly by those who were high in RWA. This effect is illustrated in Figure 3, where, from left to right, higher punishments yielded increasing social rewards from high-RWA perceivers (solid dark blue line) but tended to yield decreasing social rewards from low-RWA perceivers (dashed light blue line).

To test H2, we repeated the analysis for H1 with the authoritarian aggression subscale instead of the overall RWA scale. The result pattern was similar both in terms of statistical inference and in terms of effect sizes, $b$ = 0.04 (standardized), 95% CI [−0.01, 0.08], $p$ = .130; $b$ = 0.02 (standardized) [−0.03, 0.07], $p$ = .711; $b$ = 0.05 (standardized) [0.00, 0.10], $p$ = .038. Thus, H2 was also supported.

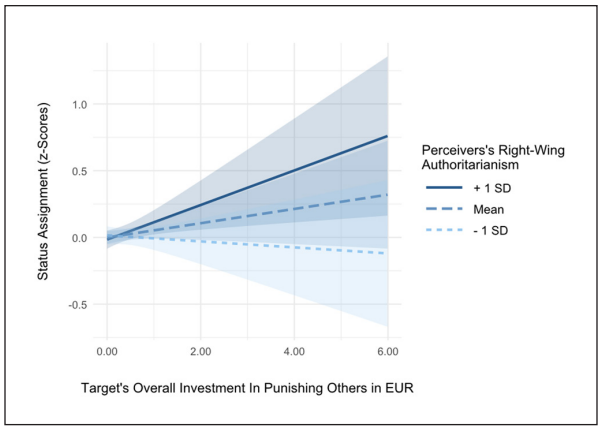



**Figure 3.** Status assignment by the target's punishment sum and the perceiver's RWA.
*Note.* RWA = right-wing authoritarianism.

### *Robustness Checks*

First, estimating a multiple-indicator multilevel model that simultaneously includes all three status indicators supports H1. The main effect of targets' punishment was significant, $b = 0.03$ (standardized), 95% CI [0.00, 0.06], $p = .033$, indicating that targets who punished others more were attributed a higher status by perceivers. Importantly, the predicted interaction between targets' punishment and perceivers' RWA also remained significant, $b = 0.04$ (standardized) [0.00, 0.07], $p = .028$. Although the coefficients were comparable to those in the main analysis, the multiple-indicator specification provides greater statistical power.

In the second robustness check, gender had a non-significant main effect ($b = -0.01$ (standardized), 95% CI [−0.04, 0.03], $p = .748$) and did not interact with punishment ($b < 0.01$ (standardized) [−0.03, 0.03], $p = .977$) or RWA ($b < 0.01$ (standardized) [−0.03, 0.03], $p = .815$). The H1 interaction remained significant when controlling for gender, $b = 0.04$ (standardized) [0.00, 0.07], $p = .033$.

Third, we tested whether the H1 targets' punishment × perceivers' RWA interaction differed across the three status indicators by allowing the predictors and their interaction to vary by indicator, $b = 0.05$ (standardized), 95% CI [0.01, 0.10], $p = .020$. However, a likelihood-ratio test indicated that allowing status indicator-specific effects did not improve model fit, $\Delta\chi^2(6) = 6.31$, $p = .389$, suggesting that the interaction did not differ significantly across the three status indicators.

Fourth, excluding the ratings of two participants who indicated that they were older than 50 despite the fact that 50 was communicated during recruitment as the maximum age did not meaningfully change the main result for targets' punishment, $b = 0.03$ (standardized), 95% CI [0.00, 0.06], $p = .039$, or the interaction of RWA × targets' punishment, $b = 0.03$ (standardized) [0.00, 0.07], $p = .033$.

Fifth, we additionally included perceivers' role (i.e., contributor vs. punisher) in the model to more broadly examine whether contributors and punishers differ in how they assign status. Status indicator-specific intercepts were included and differed from zero in this model. The main effect of perceivers' role was significant, $b = -0.10$ (standardized), 95% CI [−0.14, −0.06], $p < .001$, indicating that perceivers who could punish others themselves attributed less status to punishers compared to perceivers who were contributors. The main effect of targets' punishment remained significant, $b = 0.04$ (standardized) [0.01, 0.07], $p = .008$, and most importantly, the interaction between perceivers' RWA and targets' punishment also remained significant, $b = 0.04$ (standardized) [0.00, 0.07], $p = .024$. Compared to the model without perceivers' role and a model with an interaction of perceivers' role and the other predictors, the model including perceiver's role without interaction provided the best fit, indicating that role is important but does not moderate the critical RWA and targets' punishment interaction, $\Delta\chi^2_{\text{interaction}} = 1.59$, $p = .662$.

In the last robustness check, we included all status judgments in the dataset. The multilevel model with multiple indicators revealed a significant negative interaction between targets' punishment and perceivers' RWA (H1), $b = -0.03$ (standardized), 95% CI [−0.06, 0.00], $p = .043$. This negative interaction indicates that perceivers high in RWA attributed less additional status to targets who engaged in higher levels of punishment than perceivers low in RWA, which is the opposite pattern to what we observed in the main analysis. This suggests that, at the dyadic level, strong retaliation effects may occur, and that these dynamics may themselves be moderated by RWA. We therefore further investigate the specific dynamics between contributors and punishers in a follow-up analysis (see Supplemental Material S5). Supplemental Table S3 presents detailed results for all robustness checks.

## Discussion

Small groups maintain cooperative norms because some group members voluntarily punish freeriders at their own cost. The present study suggests that a higher stance in the group's hierarchy may compensate for these individuals' costs but that this social reward is granted foremost by conservative and authority-orientated group members (i.e., people high in RWA). Notably, we observed this effect in a large sample of participants who interacted under controlled conditions on site and were incentivized with real money. Although the effect was not robust across all the data analytic decisions we checked, it emerged in the most defensible analysis strategy, which we assume renders the overall evidence quite sound.

The present findings may offer some insights into the inconsistent picture that has emerged to date from the literature on the effects of freeriding punishment on status

attainment in small social units (Barclay, 2006; Grosz et al., 2025b; Horita, 2010; Jordan et al., 2016; Nelissen, 2008; Ozono & Watabe, 2012). Specifically, previous failures to find evidence of status benefits for punishers may have—at least in part—resulted from not including a sufficiently large portion of high-RWA participants in the respective samples. Indeed, the study by Grosz et al. (2025b) analyzed the same data as were analyzed here, and the RWA distribution was somewhat skewed because there were more low scorers than high scorers. This appears to have led to a nonsignificant main effect of punishment on status attainment both here and in Grosz et al. (2025b). The same might hold for other studies that found no status benefits for punishers. Of course, however, methodological differences between previous studies, which were not systematically addressed here—for instance, differences in how punishment and social status were operationalized—may partly account for inconsistent findings as well.

Besides contributing to the broader literature on game theory and social dilemmas, our findings add to the political psychological literature on RWA. Research has documented greater support for drastic measures among those high in RWA (Cohrs et al., 2005; Gerber & Jackson, 2013, 2015, 2017; Saunders et al., 2016), but it has not yet tested whether this tendency also manifests in small groups and when punishments are administered by laypersons rather than authorities. It also complements the personality and social psychological literature on status attainment that has often focused on who climbs up or down the social ladder (for reviews, see Bai, 2017; Grosz et al., 2025b) but less on individual differences among the group members who "push or pull" those on the ladder. Finally, our study nicely complements a finding from sexual selection research where a preference for males who helped victims rather than punished offenders in a decision game was found mostly in females with a left-wing political ideology but less in those with right-wing political views (Ferguson et al., 2019).

### *Implications and Future Directions*

The most important insight from the current experiment is that punishment of freeriding in small groups is appreciated only by some but may be dismissed by others. Thus, although freeriding punishment may successfully foster norm compliance, it may also foster ideological polarization. This has the ironic implication that, even though punishment is likely often driven by a sense of attachment to one's group, it may work against rather than toward group cohesion in the long run. Thus, while groups benefit from a *possibility* for social punishments (Fehr & Gächter, 2002), they might prosper the most when this possibility is used rather rarely and mildly.

However, more research is needed to support these conclusions and speak to applied settings. Even though the strength of the present study's design was that it afforded a large degree of control over the possible behavioral dynamics in the group, the major limitation of the present study is its ecological validity. Specifically, despite sitting in the same room, group members interacted in an entirely computer-mediated, anonymous fashion, which on the one hand forestalled unwanted influences (e.g., status assignments based on physical appearance), but on the other hand created an artificial situation. As such, it remains unclear how well the results will generalize to everyday face-to-face interactions, or to real-world social or economic situations. For example, in face-to-face settings such as travel or work groups, individuals can directly observe others' behavior and draw on a wide range of social cues (e.g., engagement, tone of voice, and nonverbal signals) when evaluating norm compliance and punishment of freeriding. Based on our data, we cannot draw conclusions about how the observed effects would translate to such settings. By contrast, the present design more closely resembles anonymous, computer-mediated interactions, such as those found in online Q&A communities.

More broadly, the present study was designed as a first step to examine whether this core psychological mechanism, that individuals high in RWA value norm enforcement through punishment, emerges under highly controlled conditions. Yet more research is needed to clarify the scope and boundary conditions of the present findings, especially in ecologically valid real-world face-to-face scenarios.

Beyond the implications of our findings for real-world scenarios, it is important to note that the mechanisms underlying why individuals high in RWA grant status to punishers remain unclear. We believe it is unlikely that high-RWA individuals base their status assignments on personal experience with the effectiveness of punishment, as participants in our paradigm lack sufficient exposure to evaluate this. Rather, we argue that the key question is whether participants plausibly believe that punishment helps uphold cooperative norms and respond accordingly by granting social status. Our theoretical argument highlights reasons to expect that individuals high in RWA interpret punishment as a signal of norm enforcement and commitment to group rules, leading them to attribute higher social status to punishers regardless of whether punishment objectively improves cooperation. We encourage future research to directly investigate these motivational mechanisms.

Finally, our sample was recruited from a highly educated and mostly female German population, which may explain why we generally observed quite high levels of cooperation and low levels of RWA. Because these variance restrictions might have attenuated the proposed effects, they may be even more pronounced in more diverse and representative samples from Western, industrialized societies. At the same time, however, it is questionable whether the effects would generalize to non-Western countries. Specifically, appraisals of peer punishment have been found to be less negative, on average, in several non-Western compared to Western countries (Eriksson et al., 2017), and it seems possible that the

degree to which this appraisal is contingent on political attitudes such as RWA varies across these countries as well.

In summary, there has been a notable lack of studies examining the types of individuals who approve or disapprove of freeriding punishers. Our findings propose that RWA could be a distinguishing factor that elucidates why interveners are such contentious figures, eliciting disdain from some and admiration from others.

## ORCID iDs

Richard Rau https://orcid.org/0000-0002-1490-8584
Luise Hönig https://orcid.org/0009-0009-3171-8537
Mitja D. Back https://orcid.org/0000-0003-2186-1558

## Ethical Considerations

The present study was approved by the institutional review board of the Faculty of Psychology and Sports Science of the University of Münster (Identifier: 2018-09-MG).

## Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the German Research Foundation (DFG; project number 407503175).

## Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Supplemental Material

Supplemental material is available online with this article. The study used data from the Group Interaction and Perception of Others (GIPO) project, which is comprehensively documented on a public repository (Grosz et al., 2025a; see https://osf.io/e3cpn). All data are publicly available at https://osf.io/e3cpn and https://osf.io/6enp4. All analysis scripts are publicly available at https://osf.io/6enp4.

## Notes

1. Some scholars have used the term *freeriders* to refer to individuals who contributed nothing to the group (e.g., Fischbacher & Gächter, 2010), while others have used it to describe individuals who contributed little or less than others (e.g., Carpenter, 2007; Gunnthorsdottir et al., 2007). In this paper, we use the term freerider to refer to individuals who contributed less (but not necessarily nothing) to the group yet received the same or higher benefits as others.
2. In one session, no usable data were recorded due to a technical problem ($n = 7$), and in three sessions, data were not usable due to invariant participant behavior (all participants contributed all their money in every round of every game, $n = 24$). Furthermore, one person participated in the laboratory experiment twice, so we deleted their data from the second participation.
3. See Supplemental Material S6 for exploratory analyses on group-level differences in RWA, punishment, and contribution.

## References

Adorno, T.W., Frenkel-Brunswik, E., Levinson, D. J., & Sanford, R. N. (1950). *The authoritarian personality*. Harper & Brothers.

Altemeyer, B. (1981). *Right-wing authoritarianism*. University of Manitoba Press.

Altemeyer, B. (1998). The other "Authoritarian Personality." *Advances in Experimental Social Psychology*, *30*, 47–92. https://doi.org/10.1016/S0065-2601(08)60382-2

Anderson, C., Hildreth, J. A. D., & Howland, L. (2015). Is the desire for status a fundamental human motive? A review of the empirical literature. *Psychological Bulletin*, *141*(3), 574–601. https://doi.org/10.1037/a0038781

Back, M. D., & Kenny, D. A. (2010). The social relations model: How to understand dyadic processes. *Social and Personality Psychology Compass*, *4*(10), 855–870. https://doi.org/10.1111/j.1751-9004.2010.00303.x

Bai, F. (2017). Beyond dominance and competence: A moral virtue theory of status attainment. *Personality and Social Psychology Review*, *21*(3), 203–227. https://doi.org/10.1177/1088868316649297

Barclay, P. (2006). Reputational benefits for altruistic punishment. *Evolution and Human Behavior*, *27*(5), 325–344. https://doi.org/10.1016/j.evolhumbehav.2006.01.003

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beierlein, C., Asbrock, F., Kauff, M., & Schmidt, P. (2014). Die Kurzskala Autoritarismus (KSA-3): Ein ökonomisches Messinstrument zur Erfassung dreier Subdimensionen autoritärer Einstellungen. Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS). https://doi.org/10.6102/zis228

Ben-Shachar, M. S., Lüdecke, D., & Makowski, D. (2020). effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software*, *5*(56), 2815.

Burnett, G., & Bonnici, L. (2003). Beyond the FAQ: Explicit and implicit norms in Usenet newsgroups. *Library & Information Science Research*, *25*(3), 333–351.

Carpenter, J. P. (2007). Punishing free-riders: How group size affects mutual monitoring and the provision of public goods. *Games and Economic Behavior*, *60*(1), 31–51. https://doi.org/10.1016/j.geb.2006.08.011

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

Cohrs, J. C., Kielmann, S., Maes, J., & Moschner, B. (2005). Effects of right-wing authoritarianism and threat from terrorism on restriction of civil liberties. *Analyses of Social Issues and Public Policy*, *5*(1), 263–276. https://doi.org/10.1111/j.1530-2415.2005.00071.x

Duckitt, J., & Sibley, C. G. (2009). A dual-process motivational model of ideology, politics, and prejudice. *Psychological Inquiry*, *20*(2–3), 98–109. https://doi.org/10.1080/10478400903028540

Dufner, M., Zimmermann, J., Kraft, L., Czarna, A. Z., Maciej, S., & Schmukle, S. C. (2024). Our versus their narcissist: Are narcissistic persons more popular in their ingroup than in a competing outgroup? *Personality Science*, *5*, 27000710241257107. https://doi.org/10.1177/27000710241257107

Eriksson, K., Andersson, P. A., & Strimling, P. (2016). Moderators of the disapproval of peer punishment. *Group Processes &*

*Intergroup Relations*, *19*(2), 152–168. https://doi.org/10.1177/1368430215583519

Eriksson, K., Strimling, P., Andersson, P. A., Aveyard, M., Brauer, M., Gritskov, V., Kiyonari, T., Kuhlman, D. M., Maitner, A. T., Manesi, Z., Molho, C., Peperkoorn, L. S., Rizwan, M., Stivers, A. W., Tian, Q., Van Lange, P. A. M., Vartanova, I., Wu, J., & Yamagishi, T. (2017). Cultural universals and cultural differences in meta-norms about peer punishment. *Management and Organization Review*, *13*(4), 851–870. https://doi.org/10.1017/mor.2017.42

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *The American Economic Review*, *90*(4), 980–994. https://doi.org/10.1257/aer.90.4.980

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*(6868), 137–140. https://doi.org/10.1038/415137a

Ferguson, E., Quigley, E., Powell, G., Stewart, L., Harrison, F., & Tallentire, H. (2019). To help or punish in the face of unfairness: Men and women prefer mutually-beneficial strategies over punishment in a sexual selection context. *Royal Society Open Science*, *6*(9), 181441. https://doi.org/10.1098/rsos.181441

Fischbacher, U., & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review*, *100*(1), 541–556. https://doi.org/10.1257/aer.100.1.541

Gerber, M. M., & Jackson, J. (2013). Retribution as revenge and retribution as just deserts. *Social Justice Research*, *26*(1), 61–80. https://doi.org/10.1007/s11211-012-0174-7

Gerber, M. M., & Jackson, J. (2015). Authority and punishment: On the ideological basis of punitive attitudes towards criminals. *Psychiatry, Psychology and Law*, *23* (1), 113–134. https://doi.org/10.1080/13218719.2015.1034060

Gerber, M. M., & Jackson, J. (2017). Justifying violence: Legitimacy, ideology and public support for police use of force. *Psychology, Crime & Law*, *23*(1), 79–95. https://doi.org/10.1080/1068316X.2016.1220556

Gordon, D. S., & Lea, S. E. G. (2016). Who punishes? The status of the punishers affects the perceived success of, and indirect benefits from, "moralistic" punishment. *Evolutionary Psychology*, *14*(3), 1474704916658042. https://doi.org/10.1177/1474704916658042

Grosz, M. P., Fuchs, J., & Back, M. D. (2025a). The GIPO data: A publicly available data set on personality and incentivized prosocial behavior. *Personality Science*, *6*, 27000710251348400.

Grosz, M. P., Thielmann, I., Krabbe, H., & Back, M. D. (2025b). When and why do individuals high in narcissistic rivalry attain social status? *European Journal of Personality*, *39*(3), 392–407. https://doi.org/10.1177/08902070241256142

Gunnthorsdottir, A., Houser, D., & McCabe, K. (2007). Disposition, history and contributions in public goods experiments. *Journal of Economic Behavior & Organization*, *62*(2), 304–315. https://doi.org/10.1016/j.jebo.2005.03.008

Henrich, J., & Boyd, R. (2001). Why people punish defectors. *Journal of Theoretical Biology*, *208*(1), 79–89. https://doi.org/10.1006/jtbi.2000.2202

Horita, Y. (2010). Punishers may be chosen as providers but not as recipients. *Letters on Evolutionary Behavioral Science*, *1*(1), 6–9. https://doi.org/10.5178/lebs.2010.2

Jordan, J. J., Hoffman, M., Bloom, P., & Rand, D. G. (2016). Third-party punishment as a costly signal of trustworthiness. *Nature*, *530*(7591), 473–476. https://doi.org/10.1038/nature16981

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. Guilford Press.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). LmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Liao, G. Y., Pham, T. T. L., Cheng, T. C. E., & Teng, C. I. (2020). How online gamers' participation fosters their team commitment: Perspective of social identity theory. *International Journal of Information Management*, *52*, 102095.

Nelissen, R. M. A. (2008). The price you pay: Cost-dependent reputation effects of altruistic punishment. *Evolution and Human Behavior*, *29*(4), 242–248. https://doi.org/10.1016/j.evolhumbehav.2008.01.001

Ozono, H., & Watabe, M. (2012). Reputational benefit of punishment: Comparison among the punisher, rewarder, and non-sanctioner. *Letters on Evolutionary Behavioral Science*, *3*(2), 21–24. https://doi.org/10.5178/lebs.2012.22

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Raihani, N. J., & Bshary, R. (2015). The reputation of punishers. *Trends in Ecology & Evolution*, *30*(2), 98–103. https://doi.org/10.1016/j.tree.2014.12.003

Redhead, D., Dhaliwal, N., & Cheng, J. T. (2021). Taking charge and stepping in: Individuals who punish are rewarded with prestige and dominance. *Social and Personality Psychology Compass*, *15*(2), e12581. https://doi.org/10.1111/spc3.12581

Saunders, B. A., Kelly, E., Cohen, N. P., & Guarino, C. (2016). Right-wing authoritarianism and social dominance orientation indirectly predict support for New York City's stop-&-frisk policy through prejudice. *Current Psychology*, *35*(1), 92–98. https://doi.org/10.1007/s12144-015-9364-4

Schönbrodt, F. D., Back, M. D., & Schmukle, S. C. (2012). TripleR: An R package for social relations analyses based on round-robin designs. *Behavior Research Methods*, *44*(2), 455–470. https://doi.org/10.3758/s13428-011-0150-4

Stephan, W. G., Ybarra, O., & Bachman, G. (1999). Prejudice toward immigrants. *Journal of Applied Social Psychology*, *29*(11), 2221–2237. https://doi.org/10.1111/j.1559-1816.1999.tb00107.x