Social Cognition, Vol. 42, No. 6, 2024, pp. 522–540
© 2024 The Guilford Press. https://doi.org/10.1521/soco.2024.42.6.522

COSTLINESS AND SOCIAL CONTEXT SHAPE PUNISHMENT AND PARTNER REJECTION DECISIONS

Elizabeth C. Szanton
University of Minnesota Twin Cities

Justin W. Martin, Katherine McAuliffe, and Liane Young
Boston College

Punishment and partner rejection are two ways an actor can respond to a target's uncooperative behavior. We examine how the normativity of the transgression in its social context and the costliness of the response to the target differentially impact these two responses. When considering underperformance, an *ambiguous transgression*, we anticipated that (a) actors would be more likely to respond when the transgression is counternormative, (b) partner rejection would be more sensitive than punishment to normativity, (c) actors would be more likely to respond when the cost to the target is low, and (d) punishment would be more sensitive than partner rejection to costliness. Across three studies (*N* = 543), we found support for all hypotheses except the fourth. Our findings suggest that normativity has a unique, dissociable impact on partner rejection, while the aversion to enacting a high-cost response does not depend upon having future interactions with the target.

Keywords: punishment, partner choice, cooperation

Theoretical and empirical work across disciplines has long examined two parallel processes that facilitate humans' large-scale cooperation with non-kin (Baumard et al., 2013): punishment and partner rejection. Given two partners—an actor and a noncooperative target—an actor engaging in punishment, also called *partner control*, inflicts a cost on the target with the pedagogical aim of amending the target's behavior. An actor engaging in partner rejection, also called *partner choice*, abandons the target in favor of a more cooperative alternative. A wealth of literature has examined how these two processes have shaped human evolution (Barclay, 2013, 2016; Baumard et al., 2013; Debove et al., 2015; Eisenbruch & Roney, 2017).

Elizabeth C. Szanton was formerly at Boston College.

Address correspondence to Elizabeth Szanton, University of Minnesota Twin Cities, Department of Psychology, 75 East River Parkway, Minneapolis, MN 55455. E-mail: elizabeth.szanton@gmail.com

Notably, these responses have distinct cognitive, social, and developmental features. Partner rejection is more sensitive to the target's intentions, while punishment is more sensitive to outcomes (Martin & Cushman, 2015); punishment, but not partner rejection, is influenced by in-group affiliation (Loustau et al., 2024); and children account for the target's intentions at a younger age when making partner rejection decisions relative to punishment decisions (Martin et al., 2022).

The present studies explore a novel set of inputs to punishment and partner rejection. While much work has examined features of the transgression (e.g., wrongfulness, Alter et al., 2007; outcome, Cushman, 2008) or of the transgressor (e.g., intentions, Martin et al., 2022; moral character, Delgado et al., 2005), we instead focus on the social context in which the transgression occurs. Specifically, how normative is the target's transgression relative to how *other* potential partners are acting? We also examine how costly the response is to the *target*, extending literature examining cost to the *actor* (e.g., Cheng et al., 2022). While past work has explored the severity of punishment alone (e.g., Egas & Reidl, 2008; Jiang et al., 2013), we manipulate costliness across both punishment and partner rejection, thus addressing a potential confound of one response as inherently more costly than the other. In investigating these two features—normativity and costliness—we aim to contribute to a greater understanding of the distinct mechanisms underlying punishment and partner rejection decisions.

Previous research on punishment has, by and large, examined responses to *unambiguous transgressions*, such as criminal acts (Carlsmith et al., 2002), patently unfair distributions in economic games (Gollwitzer & Denzler, 2009), or clear social norm violations such as consistently eating a coworker's lunch (Sarin et al., 2021). In the present study, we instead examine underperformance, which we will refer to as an *ambiguous transgression* given its uncertain and context-dependent status as an act of wrongdoing. On the one hand, underperformance might be understood as the failure to meet an obligation (Khan et al., 2023), or as a form of free-riding—that is, benefiting from a shared resource without contributing one's fair share of effort—both of which are considered morally wrong and worthy of punishment (Cubitt et al., 2011; Tomasello, 2020). On the other hand, underperformance might be viewed as merely the decision not to engage in a supererogatory act—that is, not going above and beyond—and thus, not particularly morally wrong or punishment-worthy (Khan et al., 2023). Ambiguous transgressions are not only an underexplored application of punishment and partner rejection models, but also a particularly useful case, because they make salient the features of normativity and costliness.

First, an ambiguous transgression makes the context in which the defection occurs particularly salient. Perceptions of a transgression's severity are not static but instead depend upon contextual features (such as the likelihood and severity of sanctions; Bregant et al., 2020; Depoorter & Vanneste, 2005; Mulder, 2018; Mulder et al., 2009). Faced with an ambiguous or unfamiliar transgression, an actor may be particularly likely to seek out contextual cues to determine whether it is punishment-worthy (such as the race of the transgressor; Wylie et al., 2024). The normativity of the target's transgression in its social context is thus one such cue that becomes particularly important for a transgression of uncertain severity. Following previous work spanning psychology (Radkani & Saxe, 2023) and sociology (Cooney & Burt, 2008), we expect actors to view transgressions as more wrong and punishment-worthy when they are less normative. As such, we predict that participants will be more likely to respond when the underperformance in question is counter-normative.

Second, an ambiguous transgression alters the cost calculation: In contrast to clear-cut punishment decisions, more severe responses are not necessarily more effective. Previous literature has suggested that punishment relies on a cost–benefit analysis, in which the cost to the transgressor must be sufficiently high relative to the cost to the actor (Egas & Riedl, 2008). In the case of an ambiguous transgression, however, exacting a high cost might be considered disproportionate (Strimling & Eriksson, 2014) and thus unfair (a widely acknowledged moral wrong in itself; Finkel et al., 2001; Graham et al., 2013; Haidt & Joseph, 2004). Why might a less severe punishment be advantageous? Recent scholarship suggests that punishment serves a primarily communicative function (Cushman et al., 2019; Sarin et al., 2021) rather than acting as a simple disincentive. In support of this theory, victims' satisfaction with a punishment imposed on a transgressor depends upon transgressors' understanding of why they were punished (Funk et al., 2014; Gollwitzer & Denzler, 2009; Gollwitzer et al., 2011; Molnar et al., 2020; although see Crockett et al., 2014; Marshall et al., 2021 for evidence that people will also engage in noncommunicative punishment). Furthermore, people believe that figurative punishments, which communicate disapproval but do not impose any cost, are recognizable and effective responses to wrongdoing (Sarin et al., 2021), and in some cases, people prefer low-cost punishments (Heffner & FeldmanHall, 2019; Jiang et al., 2013). As such, in the case of underperformance, we predict that participants will be more likely to respond when the response is low-cost, and thus effectively delivers a symbolic message but not a disproportionate consequence.

How will sensitivity to costliness and normativity vary based on the response in question? We anticipate that while actors will be more likely to respond when the transgression is counter-normative across both punishment and partner rejection, partner rejection will be particularly sensitive to the normativity of the transgression. If underperformance is counter-normative within a given social context, then most other potential partners in that context would outperform the current (underperforming) target. Thus, those making partner rejection decisions can make a straightforward calculation about their prospects based on the logical inference that pursuing partner rejection will improve them (Barclay, 2016). That inference does not hold for punishment: While counter-normativity may imply that the transgression is more worthy of punishment, it does not necessarily mean that enacting that punishment will improve the actor's prospects (Raihani et al., 2012).

Conversely, we anticipate that while actors will be more likely to respond when the response is less costly to the target across both punishment and partner rejection, punishment will be particularly sensitive to costliness. That is, an actor who anticipates future interactions with a target will be particularly wary of exacting a high cost for the target's ambiguous transgression because it might have negative social repercussions for the actor. For instance, the target might retaliate, thus undermining the attempt at partner control (Denant-Boemon et al., 2007; Nikiforakis, 2008; Wolff, 2012), or the actor might lose legitimacy as a punisher (Radkani & Saxe, 2023; Tsai, 2021). An actor making a partner rejection decision, on the other hand, may not want to exact an unfair consequence, but does not experience the threat of future interactions (Barclay & Raihani, 2016).

To summarize, we hypothesize that when responding to underperformance, an ambiguous transgression, actors will:

1. Be more likely to respond when the transgression is counter-normative, across both punishment and partner rejection.
2. Be more sensitive to the normativity of the transgression when enacting partner rejection relative to punishment.
3. Be more likely to respond when that response exacts a lower cost from the target across both punishment and partner rejection.
4. Be more sensitive to the costliness to the target when enacting punishment relative to partner rejection.

OVERVIEW OF STUDIES

Across three studies, we sought to explore the impact of how costly the punishment was to the target (costliness) and how common the transgression was (normativity) on punishment and partner rejection, using mixed within/between-subjects vignette paradigms. Due to significant similarities in both design and results, we present the studies together.

METHOD

PARTICIPANTS AND EXCLUSIONS

We recruited all samples from MTurk; sample sizes were preregistered and determined prior to the start of data collection.[1] Data, analytic code, preregistrations, and supplementary online materials (SOM) are available at https://osf.io/npyf9/?view_only=648250b52f3148ffb02b6c4bd4373ea2. Participants were paid $1.01 to take a 10–12-min online survey about decision making. See Table 1 for preregistered exclusion criteria across studies (exclusions by response condition and effects of exclusions reported in SOM); see Table 2 for demographics.

MATERIALS AND PROCEDURE

Study 1. We used an eight-condition, mixed within/between-subjects design. Participants were randomly assigned to punishment or partner rejection, our between-subjects variable. Our two within-subjects variables were transgression

[1] Because records were lost during staffing changes, we do not have access to a priori sample size or participant exclusions for Study 3.

TABLE 1. Participants' Exclusions for Studies 1, 2, and 3

	Not native English speaker	Reported paying little or no attention to >1 question	Rated self <6 on 7-point attention scale	Reaction time >3 *SD*s below log-transformed mean	Left survey prior to condition assignment	Multiple response conditions (experimenter error)
Study 1	6	58	49	16	48	0
Study 2	8	57	38	15	41	3
Study 3	6	51	32	9	38	4

Note. *SD* = standard deviation.

normativity (high-norm or low-norm) and response costliness (low-cost or high-cost). Thus, each participant saw a total of 12 vignettes, with three per condition: high-norm/high-cost; low-norm/high-cost; high-norm/low-cost; and low-norm/low-cost. The order of vignettes and of conditions was randomized.

Participants read second-person vignettes in which their success as an actor depends upon cooperation with a target (e.g., you are a waitress who shares tips with a busboy). In each vignette, the target is underperforming by not exerting effort that would benefit both partners (e.g., the busboy is not clearing extra tables in his downtime). We operationalized the ambiguity of the underperformance by emphasizing that "generally, around half" of other partners in the larger ecosystem (e.g., busboys in that city) underperform in the way of the target. Participants next learned of 100 new potential partners joining the workforce. In the high-normativity condition, more than half of those hired underperform in the way of the target. In the low-normativity condition, less than half of those hired underperform. Participants learned about a demerit system in which they could assign the target (who had already received two demerits) a third demerit. This demerit would cost the target a smaller (*low-cost* condition) or larger (*high-cost* condition) amount of money. In *partner rejection* trials, assigning a third demerit would mean the actor would be matched to a new target. We asked participants to rate their likelihood of engaging in the response (punishment or partner rejection), with anchors at 1 = "Definitely would not punish [the target]"/"Definitely would stay with [the target]" and 9 = "Definitely would punish [the target]"/"Definitely would not stay with [the target]." See Table 3 for a sample vignette and Table 4 for a description of each vignette.

Study 2. Reasoning that perhaps our costliness manipulation did not exact a sufficiently high cost, we boosted the costliness of both the low-cost and high-cost conditions by $2,000. Study 2 was otherwise identical to Study 1.

Study 3. To further increase the salience of costliness, we assured participants that punishment was likely to result in changed behavior from the target. We added one sentence to each of the punishment conditions that read: "Assigning [your

TABLE 2. Participant Demographics for Studies 1, 2, and 3

	Recruited *N* (*N* after exclusions)
Study 1	351 (223)
Study 2	301 (199)
Study 3	212 (121)

	Response condition	
	Punishment	**Partner rejection**
Study 1	125	98
Study 2	107	92
Study 3	63	58

	Age, years
Study 1	*M* = 37.62, *SD* = 11.37
Study 2	*M* = 39.28, *SD* = 12.95
Study 3	*M* = 37.33, *SD* = 10.99
	Social politics [Economic politics]
	1 = *strongly liberal*, 7 = *strongly conservative*
Study 1	*M* = 3.40, *SD* = 1.76 [*M* = 3.99, *SD* = 1.78]
Study 2	*M* = 3.25, *SD* = 1.87 [*M* = 3.67, *SD* = 1.83]
Study 3	*M* = 2.99, *SD* = 1.66, [*M* = 3.52, *SD* = 1.83]

	Gender		
	Men	**Women**	**“Other” or undisclosed**
Study 1	121	99	3
Study 2	106	92	1
Study 3	42	72	7

	Race/Ethnicity						
	White	**Black or African American**	**Asian or Asian American**	**Native American**	**Native Hawaiian or Pacific Islander**	**Hispanic or Latino**	**“Other” or undisclosed**
Study 1	187	21	14	2	0	15	3
Study 2	174	11	12	3	1	19	2
Study 3	97	9	6	2	0	8	3

partner] a 3rd demerit is likely to cause [him/her] to change [his/her] behavior for the better going forward.” Study 3 was otherwise identical to Study 2.

RESULTS

Analytic Strategy

We analyzed data using linear mixed-effects models. Our models included random intercepts for vignette and for subject, as well as random slopes for costliness

TABLE 3. Flowchart Through a Sample Vignette Demonstrating the 2 (Normativity: Norm vs. Counter) × 2 (Response Condition: Punishment vs. Partner Rejection) × 2 (Costliness: Low vs. High) Design

<table>
<tr><td colspan="4">Sample Vignette: Attorneys</td></tr>
<tr><td colspan="4">You are an attorney and at your firm, attorneys work with a partner. Your partner is named George. Each case requires many hours of work, and winning cases in court relates directly to how much work is put into each case. An attorney’s win rate strongly influences the bonus they earn. Often there is not enough time in a typical work week to put in the hours necessary to guarantee a win, and so attorneys often do work during their free time. Recently, even though you have been putting in extra hours, George has not been working extra hours during his free time to prepare for court. As a result, your and George’s win rate in court has dropped and your bonuses have not been as large as in the past. Generally, around half of attorneys that who work in this area do extra work during their free time.
The firm is planning on hiring 100 new attorneys tomorrow. Based on their history from their prior firm, you know that . . .</td></tr>
<tr><td colspan="4"><b>🔀 (within subjects)</b></td></tr>
<tr><td colspan="2"><b>Low normativity</b></td><td colspan="2"><b>High normativity</b></td></tr>
<tr><td colspan="2">. . . 85 of the new attorneys work extra hours during their free time to prepare cases, and 15 of the new attorneys have not worked the extra hours during their free time to prepare cases.</td><td colspan="2">. . . 15 of the new attorneys work extra hours during their free time to prepare cases, and 85 of the new attorneys have not worked the extra hours during their free time to prepare cases.</td></tr>
<tr><td colspan="4">You are considering how to respond to George’s behavior. The firm uses a demerit system. George currently has 2 demerits.</td></tr>
<tr><td colspan="4"><b>🔀 (between subjects)</b></td></tr>
<tr><td colspan="2"><b>Punishment</b></td><td colspan="2"><b>Partner rejection</b></td></tr>
<tr><td colspan="2">According to the schedule, your pairing with George will continue for the time being. Using these demerits, you can punish George, to try to change his behavior, by assigning George a 3rd demerit. [Assigning someone a 3rd demerit costs them $2,000.] And if you assign George a 3rd demerit, George will forfeit any bonus money he has earned in the prior earning period.
*Assigning George a 3rd demerit will likely create a positive change in his behavior going forward.</td><td colspan="2">Using these demerits, you can end your partnership with George. If you assign George a 3rd demerit, George’s partnership with you will end, and you can find a new partner amongst the other attorneys at the firm. [Assigning someone a 3rd demerit costs them $2,000].
And because of the time it would take George to find a new partner due to his demerits.</td></tr>
<tr><td colspan="4"><b>🔀 (within subjects)</b></td></tr>
<tr><td><b>Low cost</b></td><td><b>High cost</b></td><td><b>Low cost</b></td><td><b>High cost</b></td></tr>
<tr><td>. . . This would cost him $3,176 in reduced earnings. So, in total, assigning George a 3rd demerit will cost him $3,176 [$5,176].</td><td>. . . This would cost him $10,824 in reduced earnings. So, in total, assigning George a 3rd demerit will cost him $10,824 [$12,824].</td><td>This would end up costing George $3,176. So, in total, assigning George a 3rd demerit will cost him $3,176 [$5,176].</td><td>This would end up costing George $10,824. So, in total, assigning George a 3rd demerit will cost him $10,824 [$12,824].</td></tr>
</table>

Note. Brackets indicate wording included in Studies 2 and 3; asterisk indicates wording included in Study 3.

TABLE 4. Scenario, Participant's Role, and Partner's Underperformance by Vignette

Vignette	Scenario	Participant's role	Partner's underperformance
1	Furniture production plant	Structure worker	Cosmetic worker does not put in extra time on extra decorations
2	Tennis team	Tennis player	Tennis partner does not practice on her own
3	Police force	Police officer	Partner does not agree to patrol during lunch break
4	Law firm	Attorney	Partner does not put in extra hours to prepare for court
5	Professional pool	Pool player	Partner fails to score points for team
6	School history club	Presenter	Partner does not put in extra detail work on poster presentation
7	Badminton club	Badminton player	Partner does not practice for tournaments
8	Software company	Salesperson reaching out to clients	Partner does not reach out to new clients on her own time
9	FBI	FBI agent	Partner does not agree to stay after work to do paperwork
10	High school	High school teacher	Running mate for cafeteria committee does not help make pamphlets and posters
11	Chain restaurant	Waitress	Busboy does not clean your tables during downtime
12	Ballroom competition	Ballroom dancer	Dance partner does not agree to practice outside of official practice time

(high-cost vs. low-cost) and normativity (high-norm vs. low-norm) for subjects and for vignettes.[2] The fixed effects were response type (punishment vs. partner rejection), costliness (high-cost vs. low-cost) and normativity (high-norm vs. low-norm) and all possible interactions. In cases when this model failed to converge or had a singular fit, we followed previously established guidelines to iteratively reduce the model until it converged and no longer had a singular fit (Bates et al., 2015; Matuschek et al., 2017). The outcome variable was the likelihood of engaging in the assessed response. Our primary interest was in the two-way interactions between response condition and costliness and between response condition and normativity (although we also tested for the three-way interaction between response condition, costliness, and normativity). For each model, we report a general effect size *d* (Judd et al., 2017; see SOM).

Study 1. In alignment with our first hypothesis, there was a significant main effect of normativity ($b = -1.34$, 95% CI [−1.79, −.89], $SE = .22$, $t = -6.02$, $p < .001$, $d = .46$), such that response likelihood was greater when the transgression was

[2] Our preregistered analyses included random slopes of participants nested within vignettes. However, because each participant saw each vignette once, this is not possible, and was written in error.

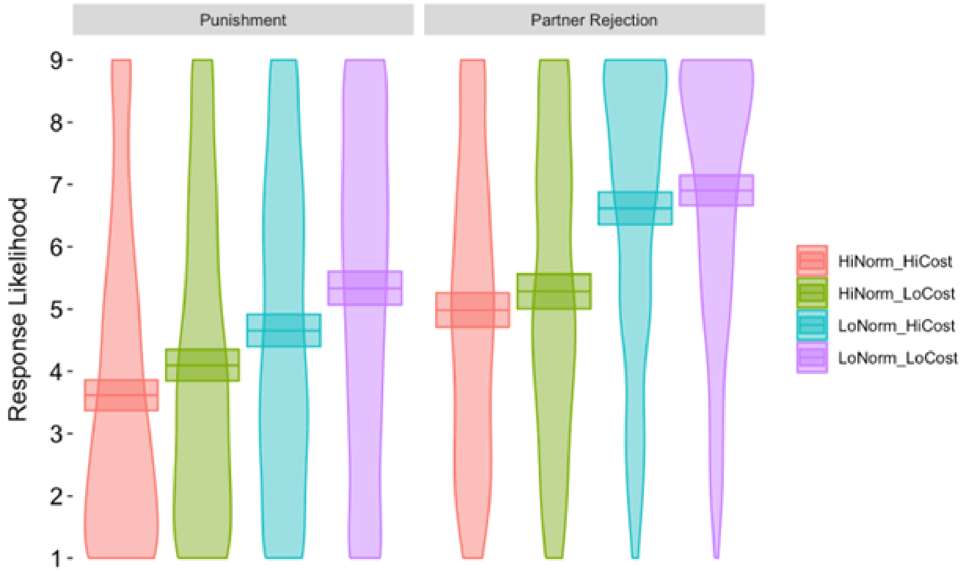

FIGURE 1. Study 1 mean response likelihood as a function of response type (punishment or partner rejection), normativity (lo-hi), and costliness (lo-hi). Center bars represent group means, and error bars represent standard error.

counter-normative. In alignment with our second hypothesis, there was a significant two-way interaction between normativity and response condition (b = −.50, 95% CI [−.92, −.08], SE = .21, t = −2.35, p = .020, d = −.17), such that normativity had a greater impact on partner rejection than on punishment. In alignment with our third hypothesis, there was a significant main effect of costliness, such that response likelihood was greater when the costliness to the target was low (b = −.44, 95% CI [−.60, −.28], SE = .08, t = −5.50, p < .001, d = .15). Contrary to our fourth hypothesis, however, costliness did not exert a significantly stronger effect on punishment than on partner rejection (b = .25, 95% CI [−.06, .57], SE = .16, t = 1.56, p = .12, d = .09). See Figure 1 for means by condition.

There was not a significant three-way interaction between costliness, normativity, and response condition (b = –.15, 95% CI [–.67, .38], SE = .27, t = –.55, p = .59, d = .05). There was a significant main effect of response condition (b = 1.48, 95% CI [1.05, 1.91], SE = .21, t = 6.90, p < .001, d = –.51), such that participants were more likely to engage in partner rejection than in punishment.

Study 2. In alignment with our first hypothesis, there was a main effect of normativity (b = −1.16, 95% CI [−1.63, −.69], SE = .23, t = −5.02, p < .001, d = .40), such that response likelihood was higher when the transgression was counter-normative. In alignment with our second hypothesis, there was a significant two-way

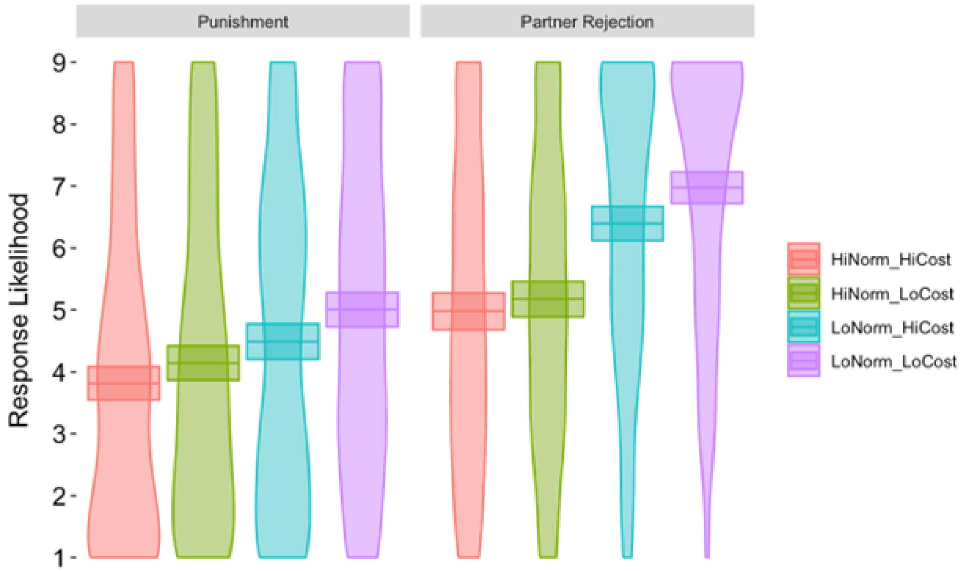

FIGURE 2. Study 2 mean response likelihood as a function of response type (punishment or partner rejection), normativity (lo-hi), and costliness (lo-hi). Center bars represent group means, and error bars represent standard error.

interaction between response condition and normativity (b = –.82, 95% CI [–1.26, –.37], SE = .23, t = –3.58, p < .001, d = –.14), such that normativity impacted response likelihood more so in the partner rejection condition than in the punishment condition. In alignment with our third hypothesis, there was a main effect of costliness (b = –.39, 95% CI [–.55, –.24], SE = .08, t = –4.98, p < .001, d = .14), such that response likelihood was higher when cost to the target was low. Contrary to our fourth hypothesis, the interaction between response condition and costliness (b = .04, 95% CI [–.27, .36], SE = .16, t = .28, p = .78, d = .01) was nonsignificant: thus, punishment was not more sensitive to costliness, relative to partner rejection. See Figure 2 for means by condition.

There was not a significant three-way interaction between response condition, costliness, and normativity on likelihood of engaging in a response (b = .04, 95% CI [–.49, .57], SE = .27, t = .15, p = .88, d = –.01). There was a significant main effect of response condition (b = 1.57, 95% CI [1.09, 2.04], SE = .24, t = 6.65, p < .001, d = –.54), such that participants were more likely to engage in partner rejection than in punishment.

Study 3. In alignment with our first hypothesis, there was a significant main effect of normativity (b = –.90, 95% CI [–1.43, –.38], SE = .26, t = –3.46, p = .004, d = .15), such that response likelihood was lower when the transgression was counter-normative. In alignment with our second hypothesis, there was a significant two-way

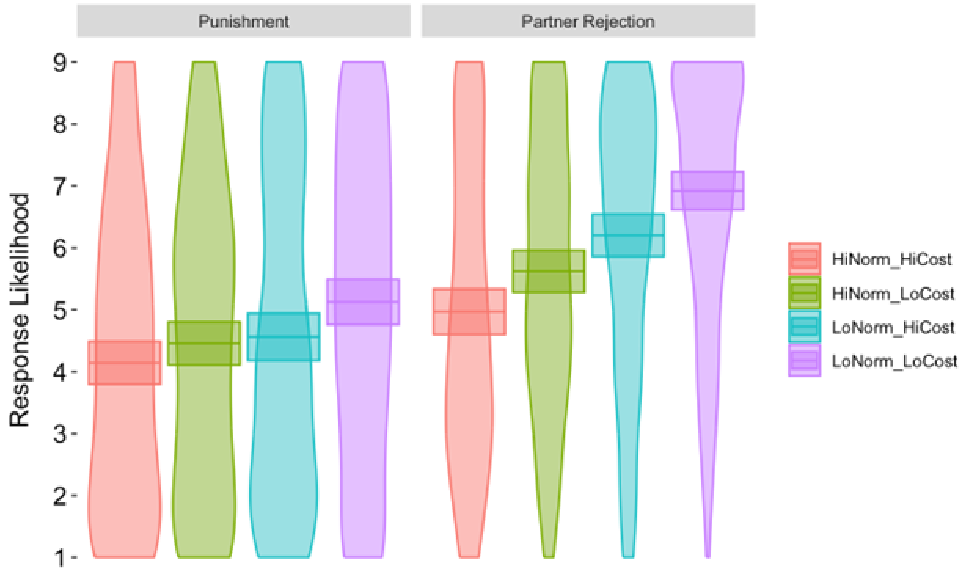

FIGURE 3. Study 3 mean response likelihood as a function of response type (punishment or partner rejection), normativity (lo-hi), and costliness (lo-hi). Center bars represent group means, and error bars represent standard error.

interaction between response condition and normativity (*b* = −.67, 95% CI [−1.19, −.15], *SE* = .27, *t* = −2.51, *p* = .013, *d* = −.11), such that normativity had a greater impact on partner rejection than on punishment. In alignment with our third hypothesis, there was a significant main effect of response costliness (*b* = −.58, 95% CI [−.80, −.37], *SE* = .11, *t* = −5.30, *p* < .001, *d* = .10), such that response likelihood was higher when costliness to the target was low. Contrary to our fourth hypothesis, there was not a significant interaction between response condition and costliness (*b* = −.17, 95% CI [−.61, .26], *SE* = .22, *t* = −.79, *p* = .43, *d* = −.03): Punishment was not more sensitive to costliness, relative to partner rejection. See Figure 3 for means by condition.

There was not a significant three-way interaction between response condition, costliness, and normativity (*b* = −.11, 95% CI [−.84, .62], *SE* = .37, *t* = −.29, *p* = .77, *d* = .02). There was a significant main effect of response condition (*b* = 1.39, 95% CI [.83, 1.95], *SE* = .28, *t* = −5.01, *p* < .001, *d* = −.23), such that participants were more likely to engage in partner rejection than in punishment.

Differences by Vignette

Full vignette effects are detailed in SOM; we describe here the patterns common across our three studies (see Table 5). Participants' overall response likelihood was higher than average for Vignette 4 (Law firm) and Vignette 8 (Sales agency); conversely, overall response likelihood was lower for Vignette 3 (Police force) and

TABLE 5. Vignette Effects Common Across Studies

	b	95% CI	*SE*	*t*	*p*	*d*
Vignette main effect: Law firm (V4)						
Study 1	.64	[.41, .86]	.12	5.46	< .001***	.23
Study 2	.85	[.61, 1.08]	.12	7.10	< .001***	.30
Study 3	.32	[.01, .64]	.16	2.02	.043*	.12
Vignette main effect: Sales agency (V8)						
Study 1	.23	[.003, .46]	.12	1.98	.048*	.08
Study 2	.31	[.07, .54]	.12	2.58	.010*	.11
Study 3	.32	[.01, .63]	.16	1.99	.046*	.12
Vignette main effect: Police force (V3)						
Study 1	−.82	[−1.05, −.59]	.12	−6.98	< .001***	−.29
Study 2	−.48	[−.71, −.24]	.12	−3.97	< .001***	−.17
Study 3	−1.09	[−1.40, −.78]	.16	−6.81	< .001***	−.41
Vignette main effect: High school (V10)						
Study 1	−.37	[−.59, −.14]	.12	−3.14	.002**	−.13
Study 2	−.41	[−.64, −.18]	.12	−3.43	< .001***	−.15
Study 3	−.43	[−.75, −.12]	.16	−2.71	.007**	−.16
Normativity × vignette interaction: Tennis team (V2)						
Study 1	−.69	[−1.14, −.24]	.23	−2.96	.003**	−.73
Study 2	−1.13	[−1.60, −.67]	.24	−4.71	< .001***	−.85
Study 3	−1.87	[−2.49, −1.25]	.32	−5.82	<.001***	−1.08
Normativity × vignette interaction: Furniture production plant (V1)						
Study 1	.93	[.48, 1.39]	.23	4.01	< .001***	−.15
Study 2	.81	[.35, 1.27]	.24	3.39	< .001***	−.14
Study 3	1.46	[.84, 2.09]	.32	4.51	< .001***	.20
Normativity × vignette interaction: Badminton club (V7)						
Study 1	.85	[.39, 1.31]	.24	3.62	< .001***	−.18
Study 2	.93	[.47, 1.40]	.24	3.90	< .001***	−.09
Study 3	1.04	[.42, 1.65]	.32	3.23	.001**	.04
Normativity × response condition × vignette interaction: Badminton club (V7)						
Study 1	1.13	[.20, 2.04]	.48	2.37	.018*	−.04
Study 2	1.74	[.82, 2.66]	.48	3.62	< .001***	.08
Study 3	1.59	[.30, 2.89]	.68	2.34	.019*	.25

Vignette 10 (High school). In addition, there were significant two-way interactions between normativity and vignette, such that Vignette 2 (Tennis team) was particularly sensitive to normativity information, while Vignette 1 (Furniture production plant) and Vignette 7 (Badminton club) were particularly insensitive to normativity. Finally, there was a significant three-way interaction between normativity, response condition, and vignette for Vignette 7, such that insensitivity of the

vignette to normativity information was particularly pronounced in the partner rejection condition.

DISCUSSION

In the present work, we examined how the normativity of the transgression in its social context and the costliness of the response to the target influenced the likelihood of engaging in punishment and partner rejection. These two inputs are particularly salient in the case of underperformance, an ambiguous transgression. In measuring how these inputs differentially influence punishment and partner rejection decisions, the present work sought to further elucidate the distinct mechanisms that drive two processes critical to the evolution and maintenance of human cooperation.

Our first hypothesis was that actors would be more likely to pursue both punishment and partner rejection when the transgression was counter-normative because these transgressions would be considered more severe. Our second hypothesis was that partner rejection would be especially sensitive to transgression normativity because when underperformance is counter-normative, a logical leap is that an alternative partner would likely perform better. Our third hypothesis was that actors would be more likely to pursue both punishment and partner rejection when the costliness of the response to the target was low because a high-cost punishment might be considered an unfair response to an ambiguous transgression. Our fourth hypothesis was that punishment would be especially sensitive to costliness because future interactions with the target would exert pressure to maintain a cooperative relationship, avoid retaliation, and maintain legitimacy in the social ecosystem. In all three studies, our results were consistent: we found support for the first three hypotheses, but not the fourth.

In support of our first hypothesis, across three studies, participants in both conditions were more likely to respond when the transgression was counter-normative. This finding suggests that, consistent with previous research (e.g., Radkani & Saxe, 2023), actors faced with an ambiguous or unfamiliar transgression use rarity as a cue about the transgression's severity, and thus, how worthy it is of punishment. We similarly found support for our second hypothesis: Across all studies, there was a significant interaction between response condition and normativity, such that partner rejection was more sensitive to transgression normativity than punishment was. As such, while punishment and partner rejection are both sensitive to transgression normativity, partner rejection appears to be unique in involving a straightforward calculation of future benefits: How will an alternative partner perform, relative to the current target? Our findings align with previous models examining the "biological market" of partner rejection, in which incentives drive actors to abandon uncooperative targets in favor of alternatives who can provide greater benefits (Barclay, 2016). By contrast, while punishment may aim to amend a partner's behavior, it involves an uncertain payoff (Raihani et al., 2012), and thus is not necessarily subject to the same analysis of future benefits.

We also found support for our third hypothesis: Across three studies, participants in both conditions were more likely to respond when the cost incurred by the target would be low. In contrast to previous literature suggesting that people prefer punishments that exact a high cost from the target (Egas & Riedl, 2008), these findings may support the communicative inference model of punishment (Cushman et al., 2019; Sarin et al., 2021), such that delivering a message is more important than inflicting a high cost. Our results also provide novel insight into ambiguous transgressions, suggesting that when faced with uncertainty about the seriousness of the transgression, actors would prefer to punish less harshly.

Regarding our fourth hypothesis, findings were indeterminate: Across all three studies, punishment was not significantly more sensitive to costliness to the target than partner rejection. Cost did not differentially impact punishment relative to partner rejection even when increasing the salience of costliness by boosting the baseline cost incurred by the target, and when making punishment more desirable by assuring actors of its efficacy. While we can make no definite claims about a null effect, we suggest some possible interpretations of this finding, should future work corroborate that both punishment and partner rejection are sensitive to costliness, but not one more than the other. Perhaps this finding suggests a domain-general reluctance to inflict a high-cost response, exemplifying harm aversion (Cushman et al., 2012; Miller & Cushman, 2013; Miller et al., 2014) or a gut-level unwillingness to hurt others. Alternatively, perhaps actors fear the reputational risk of exacting a high-cost punishment for an ambiguous transgression, even if they will not, as a feature of our design, meet the target again. Indeed, punishment that is not clearly warranted can be a reputational hazard to an actor, such as when an actor punishes a stranger for a mild norm violation (Eriksson et al., 2017), punishes an in-group member (Sun et al., 2023), exacts too severe a punishment (Strimling & Eriksson, 2014), or receives payment to punish (Rai, 2022). Further work is needed to empirically examine the merits of these potential explanations. For instance, a future study might directly measure the aversiveness of exacting a high-cost response and assess its impact on an actor's reputation in the context of an ambiguous transgression across both punishment and partner rejection trials. Relatedly, future work might examine whether an actor would be less reluctant to punish if guaranteed anonymity, or if their partner has broken a phantom rule (a rule that is frequently broken but rarely enforced; Wylie & Gantman, 2023) in addition to underperforming.

The present work's findings should be interpreted in the context of its limitations. Given our between-subjects design, we are unable to make direct comparisons between the punishment and partner rejection conditions. Future work might directly compare the likelihood of punishment and partner rejection using a fully within-subjects design. We observed variation in response likelihood across vignettes (see SOM); future designs might seek to further standardize trials to reduce noise. Our MTurk participant pool led to high numbers of excluded participants; while exclusions were preregistered and did not impact any primary outcomes (see SOM), future studies might use platforms found to yield higher-quality human subjects data, such as Prolific (Chmielewski & Kucker, 2020; Douglas et al., 2023). Despite shortcomings, the present studies contribute novel insight into the inputs that shape cooperative behavior, demonstrating that punishment and partner rejection are both more likely when the transgression is counter-normative and when the cost incurred by the target is low. Transgression normativity has a specific, dissociable influence on partner rejection, which depends upon a strategic calculation of future benefits. By contrast, the preference for low-cost responses across conditions may suggest an aversion to severe sanctions for ambiguous transgressions regardless of whether the actor will experience future interactions with the target.

Acknowledgments. We would like to thank Kayla Carew for her work as a research assistant on this project, and Nathan Liang for his assistance in data analysis. Support for this work was provided by the John Templeton Foundation (grant # 61061) and the Boston College Virtue Project.

Declaration of interest. The authors declare no conflicts of interest.

Data availability. Data, analytic code, preregistrations, and supplementary online materials (SOM) can be found at https://osf.io/npyf9/?view_only=648250b52f3148ffb02b6c4bd4373ea2.

REFERENCES

Alter, A. L., Kernochan, J., & Darley, J. M. (2007). Transgression wrongfulness outweighs its harmfulness as a determinant of sentence severity. *Law and Human Behavior*, *31*, 319–335. https://doi.org/10.1007/s10979-006-9060-x

Barclay, P. (2013). Strategies for cooperation in biological markets, especially for humans. *Evolution and Human Behavior, 34*(3), 164–175. https://doi.org/10.1016/j.evolhumbehav.2013.02.002

Barclay, P. (2016). Biological markets and the effects of partner choice on cooperation and friendship. *Current Opinion in Psychology, 7*, 33–38. https://doi.org/10.1016/j.copsyc.2015.07.012

Barclay, P., & Raihani, N. (2016). Partner choice versus punishment in human prisoner's dilemmas. *Evolution and Human Behavior, 37*(4), 263–271. https://doi.org/10.1016/j.evolhumbehav.2015.12.004

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious mixed models*. arXiv. https://doi.org/10.48550/arXiv.1506.04967

Baumard, N., André, J. B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences*, *36*(1), 59–78. https://doi.org/10.1017/S0140525X11002202

Bregant, J., Caruso, E. M., & Shaw, A. (2020). Crime because punishment? The inferential psychology of morality and punishment. *University of Illinois Law Review, 2020*(4), 1177–1207.

Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology, 83*(2), 284–299. https://doi.org/10.1037/0022-3514.83.2.284

Cheng, X., Zheng, L., Liu, Z., Ling, X., Wang, X., Ouyang, H., Chen, X., Huang, D., & Guo, X. (2022). Punishment cost affects third parties' behavioral and neural responses to unfairness. *International Journal of Psychophysiology*, *177*, 27–33. https://doi.org/10.1016/j.ijpsycho.2022.04.003

Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science*, *11*(4), 464–473. https://doi.org/10.1177/1948550619875149

Cooney, M., & Burt, C. H. (2008). Less crime, more punishment. *American Journal of Sociology*, *114*(2), 491–527. https://doi.org/10.1086/592425

Crockett, M. J., Özdemir, Y., & Fehr, E. (2014). The value of vengeance and the demand for deterrence. *Journal of Experimental Psychology: General*, *143*(6), 2279–2286. https://doi.org/10.1037/xge0000018

Cubitt, R. P., Drouvelis, M., & Gächter, S. (2011). Framing and free riding: Emotional responses and punishment in social dilemma games. *Experimental Economics*, *14*, 254–272. https://doi.org/10.1007/s10683-010-9266-0

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*(2), 353–380. https://doi.org/10.1016/j.cognition.2008.03.006

Cushman, F., Gray, K., Gaffey, A., & Mendes, W. B. (2012). Simulating murder: The aversion to harmful action. *Emotion*, *12*(1), 2–7. https://doi.org/10.1037/a0025071

Cushman, F., Sarin, A., & Ho, M. (2019). Punishment as communication. In *The Oxford handbook of moral psychology* (pp. 197–209). Oxford University Press.

Debove, S., André, J. B., & Baumard, N. (2015). Partner choice creates fairness in humans. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1808), 20150392. https://doi.org/10.1098/rspb.2015.0392

Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience*, *8*(11), 1611–1618. https://doi.org/10.1038/nn1575

Denant-Boemont, L., Masclet, D., & Noussair, C. N. (2007). Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Economic Theory*, *33*, 145–167. https://doi.org/10.1007/s00199-007-0212-0

Depoorter, B., & Vanneste, S. (2005). Norms and enforcement: The case against copyright litigation. *Oregon Law Review*, *84*, 1127.

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLoS One*, *18*(3), e0279720. https://doi.org/10.1371/journal.pone.0279720

Egas, M., & Riedl, A. (2008). The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1637), 871–878. https://doi.org/10.1098/rspb.2007.1558

Eisenbruch, A. B., & Roney, J. R. (2017). The skillful and the stingy: Partner choice decisions and fairness intuitions suggest human adaptation for a biological market of cooperators. *Evolutionary Psychological Science*, *3*, 364–378. https://doi.org/10.1007/s40806-017-0107-7

Eriksson, K., Andersson, P. A., & Strimling, P. (2017). When is it appropriate to reprimand a norm violation? The roles of anger, behavioral consequences, violation severity, and social distance. *Judgment and Decision Making*, *12*(4), 396–407. https://doi.org/10.1017/S1930297500006264

Finkel, N. J., Harre, R., & Lopez, J. L. R. (2001). Commonsense morality across cultures: Notions of fairness, justice, honor and equity. *Discourse Studies*, *3*(1), 5–27. https://doi.org/10.1177/1461445601003001001

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), 364–374. https://doi.org/10.1002/ejsp.782

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*, *47*, 55–130. https://doi.org/10.1016/B978-0-12-407236-7.00002-4

Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, *133*(4), 55–66. https://www.jstor.org/stable/20027945

Heffner, J., & FeldmanHall, O. (2019). Why we don’t always punish: Preferences for non-punitive responses to moral violations. *Scientific Reports*, *9*(1), 13219. https://doi.org/10.1038/s41598-019-49680-2

Jiang, L. L., Perc, M., & Szolnoki, A. (2013). If cooperation is likely, punish mildly: Insights from economic experiments based on the snowdrift game. *PLoS One*, *8*(5), e64677. https://doi.org/10.1371/journal.pone.0064677

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*, *68*, 601–625. https://doi.org/10.1146/annurev-psych-122414-033702

Khan, U., Jaffer-Diaz, M., Najafizadeh, A., & Starmans, C. (2023). Going above and beyond? Early reasoning about which moral acts are best. *Cognition*, *236*, 105444. https://doi.org/10.1016/j.cognition.2023.105444

Loustau, T., Glassman, J., Martin, J. W., Young, L., & McAuliffe, K. (2024). The impact of group membership on punishment versus partner rejection. *Scientific Reports*, *14*, 22238. https://doi.org/10.1038/s41598-024-69206-9

Marshall, J., Yudkin, D. A. & Crockett, M. J. (2021). Children punish third parties to satisfy both consequentialist and retributive motives. *Nature Human Behavior, 5*, 361–368. https://doi.org/10.1038/s41562-020-00975-9

Martin, J. W., & Cushman, F. (2015). To punish or to leave: Distinct cognitive processes underlie partner control and partner choice behaviors. *PLoS One, 10*(4), e0125193.https://doi.org/10.1371/journal.pone.0125193

Martin, J. W., Leddy, K., Young, L., & McAuliffe, K. (2022). An earlier role for intent in children’s partner choice versus punishment. *Journal of Experimental Psychology: General*, *151*(3), 597–612. https://doi.org/10.1037/xge0001093

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Miller, R., & Cushman, F. (2013). Aversive for me, wrong for you: First-person behavioral aversions underlie the moral condemnation of harm. *Social and Personality Psychology Compass*, *7*(10), 707–718. https://doi.org/10.1111/spc3.12066

Miller, R. M., Hannikainen, I. A., & Cushman, F. A. (2014). Bad actions or bad outcomes? Differentiating affective contributions to the moral condemnation of harm. *Emotion*, *14*(3), 573–587. https://psycnet.apa.org/doi/10.1037/a0035361

Molnar, A., Chaudhry, S., & Loewenstein, G. F. (2020). “It’s not about the money. It’s about sending a message!”: Unpacking the components of revenge [CESifo Working Paper No. 8102]. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3541450

Mulder, L. B. (2018). When sanctions convey moral norms. *European Journal of Law and Economics*, *46*, 331–342. https://doi.org/10.1007/s10657-016-9532-5

Mulder, L. B., Verboon, P., & De Cremer, D. (2009). Sanctions and moral judgments: The moderating effect of sanction severity and trust in authorities. *European Journal of Social Psychology*, *39*(2), 255–269. https://doi.org/10.1002/ejsp.506

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, *92*(1–2), 91–112. https://doi.org/10.1016/j.jpubeco.2007.04.008

Radkani, S., & Saxe, R. (2023). What people learn from punishment: Joint inference of wrongness and punisher’s motivations from observations of punitive choices. *Proceedings of the Annual Meeting of the Cognitive Science Society, 45*, 1027–1034. https://escholarship.org/uc/item/4hb9v0pq

Rai, T. S. (2022). Material benefits crowd out moralistic punishment. *Psychological Science*, *33*(5), 789–797. https://doi.org/10.1177/09567976211054786

Raihani, N. J., Thornton, A., & Bshary, R. (2012). Punishment and cooperation in nature. *Trends in Ecology & Evolution*, *27*(5), 288–295. https://doi.org/10.1016/j.tree.2011.12.004

Sarin, A., Ho, M. K., Martin, J. W., & Cushman, F. A. (2021). Punishment is organized around principles of communicative inference. *Cognition*, *208*, 104544. https://doi.org/10.1016/j.cognition.2020.104544

Strimling, P., & Eriksson, K. (2014). Regulating the regulation: Norms about punishment. In P. A. M. Van Lange, B. Rockenbach, & T. Yamagishi (Eds.), *Reward and punishment in social dilemmas* (pp. 52–69). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199300730.003.0004

Sun, B., Jin, L., Yue, G., & Ren, Z. (2023). Is a punisher always trustworthy? In-group punishment reduces trust. *Current Psychology*, *42*(26), 22965–22975. https://doi.org/10.1007/s12144-022-03395-2

Tomasello, M. (2020). The moral psychology of obligation. *Behavioral and Brain Sciences*, *43*, e56. https://doi.org/10.1017/S0140525X19001742

Tsai, L. L. (2021). *When people want punishment: Retributive justice and the puzzle of authoritarian popularity*. Cambridge University Press.

Wolff, I. (2012). Retaliation and the role for punishment in the evolution of cooperation. *Journal of Theoretical Biology*, *315*, 128–138. https://doi.org/10.1016/j.jtbi.2012.09.012

Wylie, J., & Gantman, A. (2023). Doesn't everybody jaywalk? On codified rules that are seldom followed and selectively punished. *Cognition*, *231*, 105323. https://doi.org/10.1016/j.cognition.2022.105323

Wylie, J., Milless, K. L., Sciarappo, J., & Gantman, A. (2024). The biased enforcement of rarely followed rules. *Personality and Social Psychology Bulletin*, 01461672241252853. https://doi.org/10.1177/01461672241252853