Supplementary material

Asymmetry and symmetry of acts and omissions in punishment, norms, and judged causality

Relationship between studies

In this supplement, we report supplementary results to the two studies reported in the main text. We also present the results of two preliminary studies which informed the design of our main studies. All four experiments studied paradigms in which subjects played an asymmetric public goods game, with punishment. The differences related to the causal structure of the groups, the mechanism by which punishment decisions were made, and whether participants gave their answers hypothetically, in advance of knowing their particular roles in the group (the "strategy" method) or made decisions live, having learned their specific role. Table S1 summarizes the differences.

Table S1. Summary of differences between studies.

Study	**Number of "actors"**	**Number of "omitters"**	**Punishment decision**	**Strategy method or "live" decisions**	**Marginal cost of punishing additional transgressors**	**Preregistration**
Preliminary 1	1	1	Simultaneous	Strategy method	Zero	https://osf.io/8fxuc
Preliminary 2	1	1	Sequential	Strategy method	Zero	https://osf.io/h5mpa
Study 1	2 or 1	1 or 2	Sequential	Live (plus one round with strategy)	Constant, positive	https://osf.io/2prkv
Study 2	2	2	Sequential	Live	Constant, positive	n/a

Study 1, additional results

Robustness check – main results including comprehension failures

The main regressions from Study 1 are recalculated in Table S2, this time including groups which failed our pre-registered comprehension criterion. The same pattern of results is observed.

Table S2. Regression models run on Study 1, with excluded data now included as additional robustness check, as per preregistration.

	(1)	(2)	(3)	(4)	(5)	(6)
Action (not omission)	1.170***	1.088***	0.951*	1.111***	1.088***	1.088***
	(0.243)	(0.246)	(0.408)	(0.258)	(0.257)	(0.257)
Jointly responsible (not solely)	-0.516*	-0.625*	-0.735	-0.645*	-0.629*	-0.629*
	(0.243)	(0.246)	(0.408)	(0.258)	(0.257)	(0.257)
Treatment order	0.438	0.309	0.451	0.153	0.152	0.043
	(0.535)	(0.549)	(0.536)	(0.577)	(0.576)	(0.571)
Action x Jointly responsible			0.438			
			(0.657)			
Asymmetry of fairness judgments				-0.132	-0.417	-0.374
				(0.288)	(0.315)	(0.313)
Action x Asymmetry of fairness judgments					0.569*	0.569*
					(0.254)	(0.254)
Additional covariates	No	Yes	No	No	No	Yes
N	592	566	592	536	536	536

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Robustness check – Strategy round results

As a further check on our results, on one round we collected participant responses via the strategy method, which involves asking for hypothetical decisions whether to press each variety of button, and what punishment policy to endorse, prior to knowing what role will be occupied on that round. Because the combinatorial space of possible options is very large, we presented subjects with a simplified set of punishment options: punish actions maximally, punish omissions maximally, punish both maximally, or punish neither. The results are summarized in Table S3. The results are broadly consistent with those observed on the ordinary rounds; actions are punished more than omissions, but there is a significant

amount of punishment of omissions. Comparing across treatments, jointly responsible behaviors (acts or omissions) are punished less frequently than solely responsible behaviors, suggesting some sensitivity on the part of participants to the efficiency consideration.

Table S3. Distribution of punishment responses at individual level in the strategy rounds for each treatment.

Treatment	Punish actions only	Punish omissions only	Punish both	Punish neither
2A (n = 210)	18.6%	11.9%	27.1%	42.4%
2O (n = 210)	24.8%	4.3%	27.6%	43.3%

Testing for the significance of these differences at the group level, we find that acts are punished significantly more frequently than omissions (Table S4). Jointly responsible behaviors are punished less than solely responsible behaviors, but the difference is not statistically significant. Consistent with the non-strategy rounds, we found a large proportion of participants were willing to punish omitters. In 2O, 31.9% proposed to punish either acts and omissions or omissions alone (95%CI: .276–.365), and in 2A the corresponding proportion was 39.0% (95%CI: .345–.438). If we focus on participants who chose to propose any punishment at all, more than half proposed to punish omitters. Broken down by treatment: in 2O, 56% (CI: .499–.625), and in 2A, 67.8% (CI: .616–.734) of punishers proposed to punish omitters.

Table S4. Frequency of subjects' decisions to punish various behavior types on the strategy round, conditional on the assumption they will be in the role of bystander. For both jointly responsible and solely responsible behaviors, acts are more likely to be punished more frequently than omissions, though the result is not statistically significant for joint (Joint: $z = 1.815$, $p = 0.0695$; Sole: $z = 2.492$, $p = 0.0122$, Wilcoxon ranksum). For both acts and omissions, the difference in punishment frequencies between jointly responsible and solely responsible behaviors is not significant (Acts: $z = 1.593$, $p = .111$; Omissions: $z = 1.087$, $p = .277$, Wilcoxon ranksum). All tests conducted at the group level, $n = 37$.

	Act	Omission
Jointly responsible	0.43	0.32
Solely responsible	0.53	0.38

Inequality

Do people perform harmful actions or omissions more frequently when they have a disadvantageous endowment at the beginning? In short, yes, see Figure S1. As subjects have more disadvantageous endowments (left side of figures), they are more likely to press red

and to refuse to press green. We also see that, in general, subjects are more likely to press red in the 2A treatment and to omit to press green in the 2O treatment, which is what we would expect, because these are conditions where it may be possible to perform the relevant act/omission without contributing to harm.

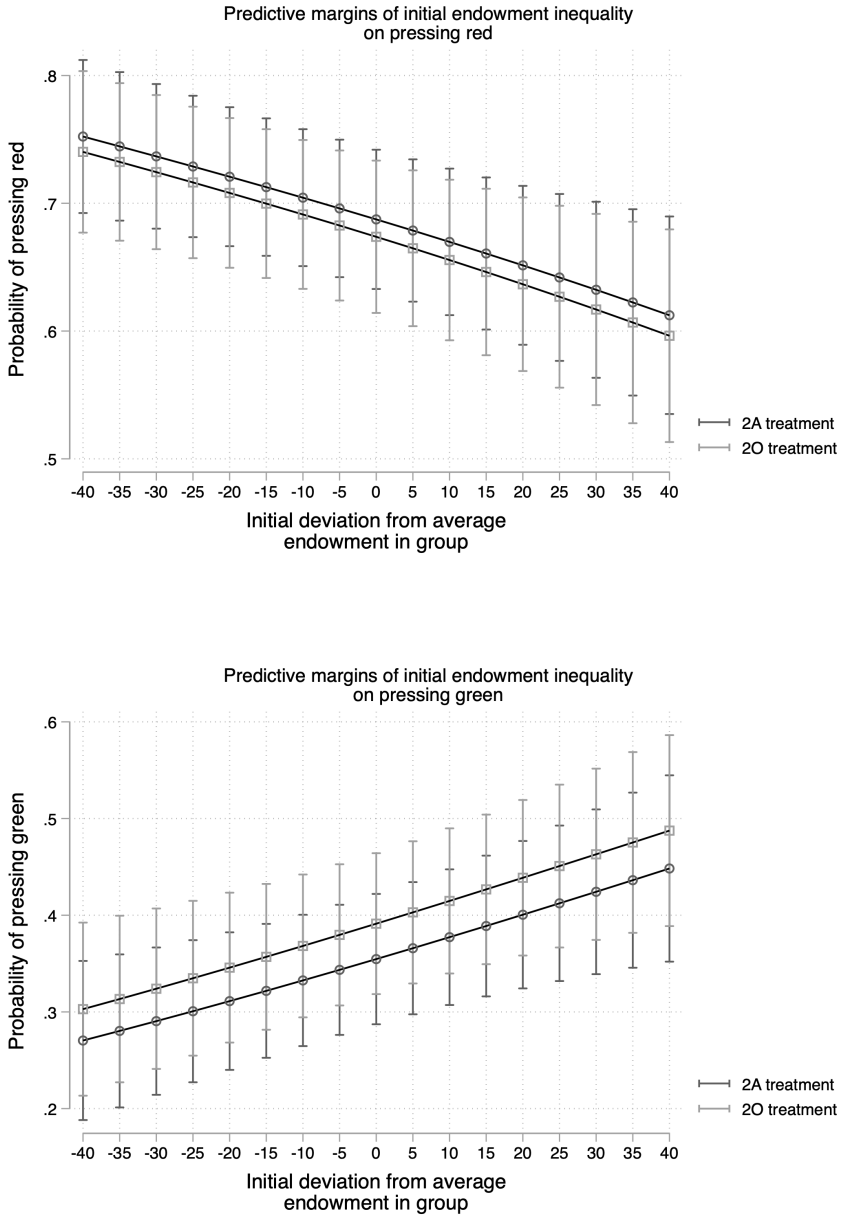


Figure S1. Logistic regression model estimated probability subject will press red/green, conditional on treatment and conditional on deviation of subject's initial endowment from the group average. Error bars indicate 95% confidence intervals.

Actual punishment dispensed

Because actual punishment requires approval of the victim, the punishment proposal data provides a richer set of observations of the group's punishment intentions. In the main text, therefore, our analyses focus only on proposed punishment. For completeness, in Tables S5 and S6, we describe the distribution of punishment proposals that are actually implemented.

Table S5: Summary of actual punishment outcomes, for rounds where harm occurred.

Treatment	Punish actions only	Punish omissions only	Punish both	Punish neither
2A (n = 163)	7%	2%	23%	67%
2O (n = 106)	8%	3%	22%	67%

Table S6. Summary distribution of mean punishment levels implemented for jointly/solely responsible acts and omissions, conditional on any punishment being proposed. Standard deviations calculated at group level.

	Actions	Omissions
Jointly responsible	2.92 (2.96)	1.43 (2.51)
Solely responsible	1.93 (2.61)	2.37 (2.96)

Punishment in no-harm scenarios

Our study also enabled us to collect data on subjects' decisions to punish others when no harm actually eventuated. In the 2A treatment, if only one person presses a red button, no harm will eventuate, but subjects still had the option of punishing that individual. And in the 2O treatment, if the red button is pressed and only one green button is pressed, then subjects could punish both the red button presser and the green refuser. We did not have any prior hypotheses regarding these scenarios. We present a summary of the results below.

First, note a problem with the comparability of punishment decisions in the 2A and 2O "no-harm" scenarios. In 2A, subjects were only able to punish one individual: the sole red presser, whereas in 2O, they had the opportunity to punish two individuals. This meant they faced a substantially different option set, and it is invidious to compare punishment expenditures across those scenarios without further information regarding the impact of the

changed option set itself. That said, we may readily compare within subjects how much they spent on punishing actors and omitters within a single scenario: in the 2O treatment, this condition is met – there is one potentially harmful action, and one potentially harmful omission – and we observed that subjects spent significantly more on punishing actors than omitters, consistent with the pattern of results in our main hypothesis (see Figure S2).

Figure S2. Histogram of punishment levels proposed for actions and omissions in the 2O treatment, where no harm actually eventuates.

As the histogram indicates, and statistical test confirms, punishment of actors was significantly higher than of omitters (Wilcoxon signed rank test, $z = 3.130$, $p = 0.002$, $n = 37$ groups).

As noted above, it is not possible to make any confident comparison of punishment of actors across the 2A and 2O treatments in no-harm scenarios, given the disanalogous option sets subjects faced. But for what it's worth, the mean amount of punishment was broadly similar – though the distribution was more extreme in the 2O treatment (see Figure S3). The distribution was not very surprising on the null hypothesis that there is no difference (Mann–Whitney test, $z = 0.851$, $p = 0.3948$).

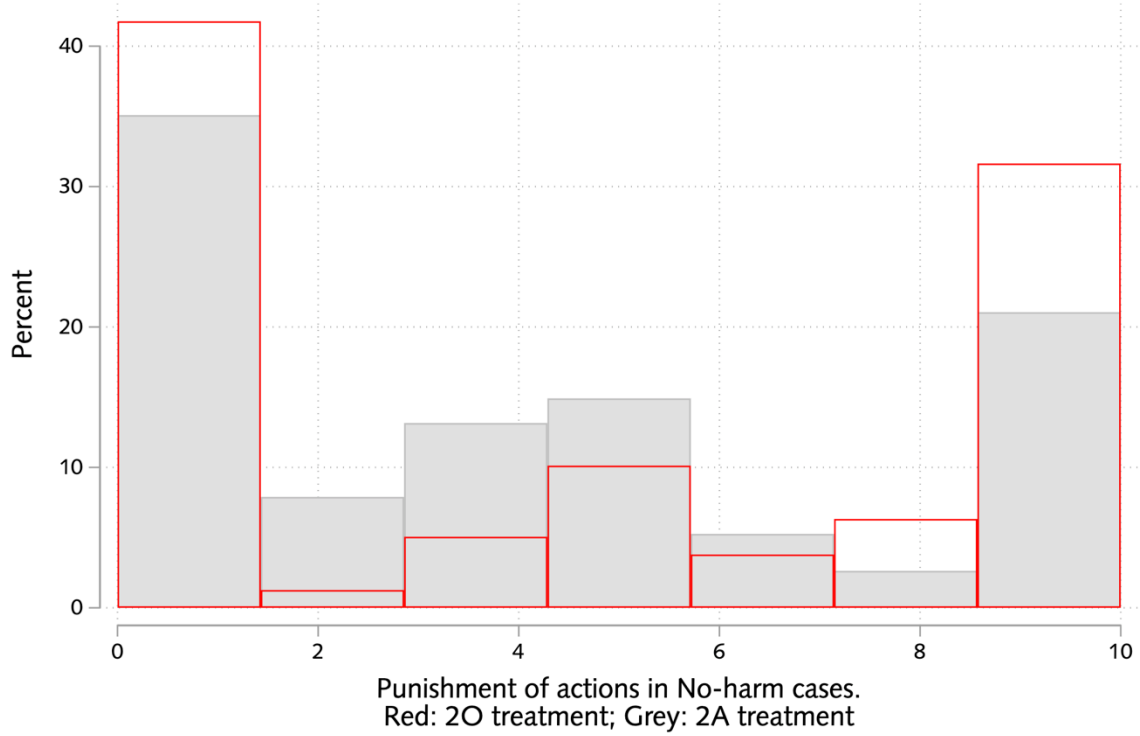


Figure S3. Histogram of proposed punishments of actions in no harm cases, across 2O and 2A treatments.

Punishment of omitters, participant level observations

Table S7. Frequency of individuals who had the opportunity to propose punishment to both actors and omitters, who actually proposed punishment of omitters, across Studies 1 and 2, with confidence intervals.

Treatment	**P(Pun omit, conditional on Pun anyone)**	**95% CI**	**P(Pun omit)**	**95% CI**
2A	51/64 (.80)	.678–.887	51/101 (.50)	.407–.602
2O	35/46 (.76)	.612–.874	35/78 (.45)	.342–.562
Combined 2A + 2O	86/110 (.78)	.694–.850	86/179 (.48)	.408–.554
2R2G	22/33 (.67)	.485–.809	22/46 (.48)	.336–.624
All combined	108/143 (.76)	.677–.812	108/225 (.48)	.415–.546

Study 2, supplementary results

No harm situations in the 2A2O study

As in the 2O treatment of Study 1, this paradigm allows for a situation where there are both doers and allowers who have behaved in a way that risked harm, without any harm actually eventuating. That is, if two people press red, but only one presses green, then both the red pressers and the green refuser risked harm, but no loss would be experienced by the victim. We compared proposed amounts of punishment for actions and omissions in this situation (see Figure S4 for histogram) and found a modest difference such that actors were punished more than omitters. (Mean per capita proposed punishments of 3.86 and 2.79 units, respectively; Wilcoxon matched pairs signrank test, $z = 1.972$, $p = 0.049$, $n = 21$ groups). Although the difference is modest, it is noteworthy that every unit of per capita punishment of doers in this scenario is twice as costly to the punisher as punishment of an omitter, because there are two red button pressers, but only one green button refuser. This might account for the difference being smaller in this case than in the analogous case from Study 1.

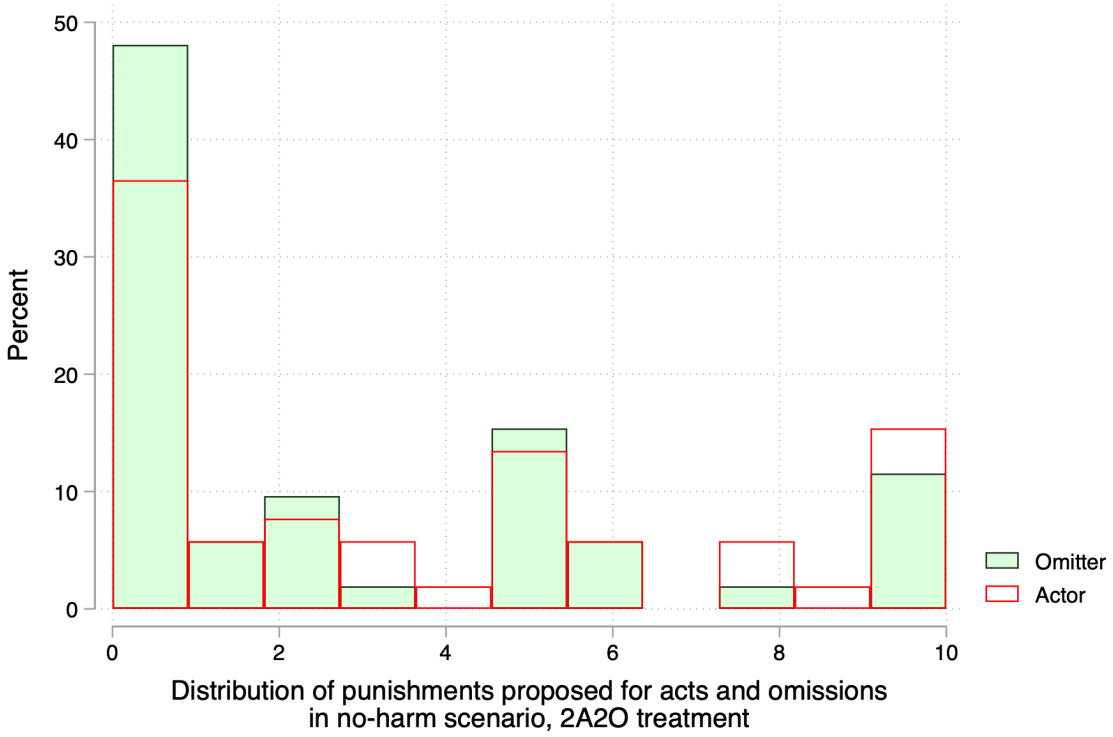


Figure S4. Histogram of proposed punishments of actions and omissions in the no harm scenario, 2O2A treatment.

Preliminary Study 1

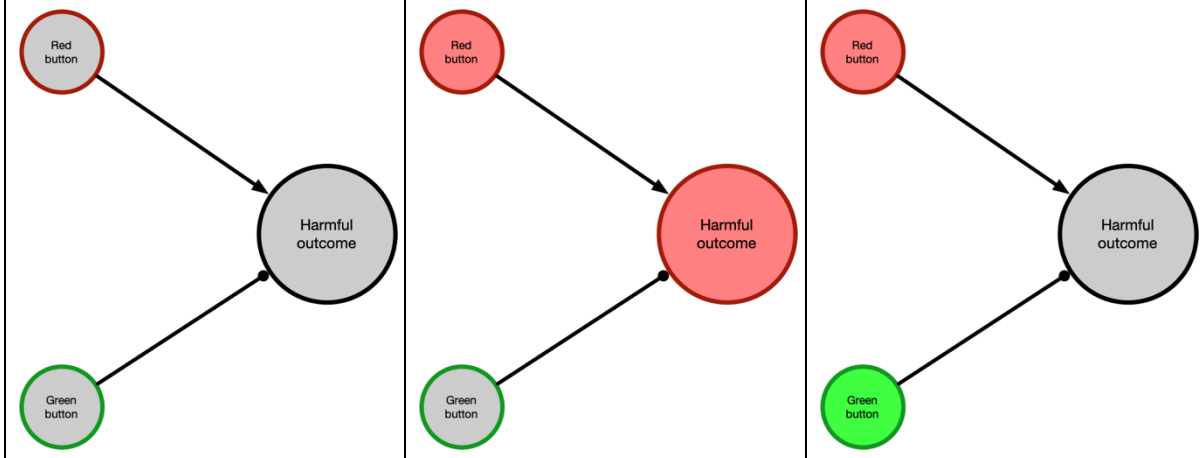
Our first Preliminary Study, like Study 1, involves a finitely repeated taking game with punishment, in fixed groups. The principal differences were that:

1. There was only one red button and one green button, so it was equally efficient to punish either actors or omitters. See Figure S5 for an overview.

2. We used the strategy method (Selten 1967) to collect players' decisions: we ask all players to decide, before they know which role they will occupy, what choice they will make, contingent on all other possible choices. We then randomly allocate subjects to roles and advise participants of their role assignment and the outcome, using the earlier responses. This enables collection of maximum possible data on individual decisions at some loss to psychological realism.

3. Punishment decisions were made by the unaffected participant and the victim simultaneously, meaning that coordination was particularly difficult to achieve.

4. There was a flate rate cost for punishing others, regardless of whether one or two people would be punished. In other words, the marginal cost of punishing an additional person was zero.



If no red button is pressed harm does not occur.

If red button is pressed and green button not pressed, harm occurs

If green button is pressed, harm is prevented.

Figure S5. Causal links between acts, omissions, and harm in the experimental setting.

We hypothesized that where subjects could choose to support an action rule (punishing actors only), an outcome rule (punish both actors and omitters equally), or an omission rule

(punish omitters only), there would be more support for the action rule than for either of the other two rules.

Methods

Participants played a repeated game for 10 rounds in fixed groups. Players were identified by anonymized labels, which were shuffled to prevent reidentification between rounds. On each round, each participant was endowed with 10 experimental currency units (ECU). At the end of the experiment one round was chosen randomly for payment, and each ECU was converted to AUD2.00. Each round, players are randomly assigned to the roles of:

- Red button presser. This player can take from a random other player, gaining 2 ECU, and reducing the other's holdings by 6 ECU.
- Green button presser. This player can restore the loss of 6 to another player, if the red button was pressed, at cost 2 to whomever presses.
- Red button victim. This player suffers a loss of 6 if the red button is pressed and is restored to 10 if the green button is pressed.
- Unaffected individual. The red button victim and the unaffected individual have the opportunity to support a deduction policy, at cost of 1 ECU each. If both players support the same deduction policy, then 6 points are deducted from whomever is specified by the relevant policy. The policies that could be chosen were:
    - Action rule (deduct from the red presser only)
    - Omission rule (deduct from the green non-presser only)
    - Outcome rule (deduct from both the red presser and the green non-presser)

In order to collect data on subjects' expectations regarding other participants' behavior, one round was accompanied by an unexpected survey task. Participants were asked to report what they believed participants would elect to do, in all roles, on the next round. To incentivize responses to this task, participants could earn a bonus payment if their answer to a given question was the most commonly given answer in the group. This method is an incentive compatible way of eliciting beliefs about norms, which in this instance we thought were likely to be influential in explaining behavior (Krupka and Weber 2013).

We also surveyed the subjects after the experiment to collect demographic information and to elicit beliefs about the fairness of being punished for pressing red/refusing to press green, how angry they would be at red-pressers and green refusers, and whether they would be likely to retaliate if they were punished for red pressing or green-refusal.

156 participants were recruited from the Monash Laboratory for Experimental Economics subject pool (62 female, 94 male; mean age 21.6 years, sd = 3.2). The experimental setup was programmed using oTree (Chen, Schonger, and Wickens 2016). Subjects were paid in cash, in private, immediately after the experiment. They earned a $10 show up fee, plus additional

earnings contingent upon their decisions in the strategic game. Average earnings were AUD31.00 (sd = 3.82).

Results

Our pre-registered hypothesis, that there would be more support for the action rule over all other rules, was not supported. While there was very little support for the omission rule, the levels of support for the action rule and the outcome rule were very similar. See Figures S6 and S7. (Action vs Outcome, $z = -0.098$, $p = 0.9217$; Action vs Omission, $z = 4.753$, $p < 0.0001$; Outcome vs Omission, $z = 5.021$, $p < 0.0001$; Wilcoxon signrank tests, all conducted at the group level.)

As a more direct test of whether there is a preference to punish acts over omissions, we compared aggregate support for the two rules that include punishing actions (act rule and outcome rule) to aggregate support for the two rules that involve punishing omissions (outcome rule and omission rule). Mean level of support for punishing actions was 0.54 of group members (sd = 0.23) and mean level of support for punishing omissions was 0.31 of group members (sd = 0.20). The difference is significant (Wilcoxon signrank test, $z = 4.76$, $p < 0.0001$).

As we noted in our pre-registration, we compared behavior in rounds 5–9 to allow for a learning period and also to exclude an end round effect. As a robustness check, we ran the same tests using data from all rounds and the same pattern was observed.

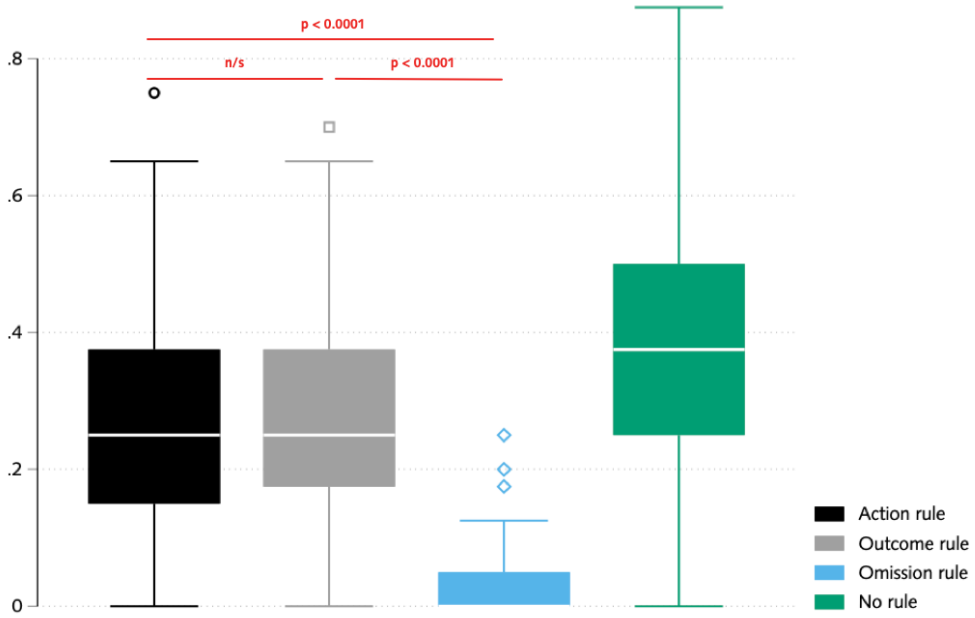


Figure S6—Levels of support within each group for the three possible rules, data from rounds 5–9 only

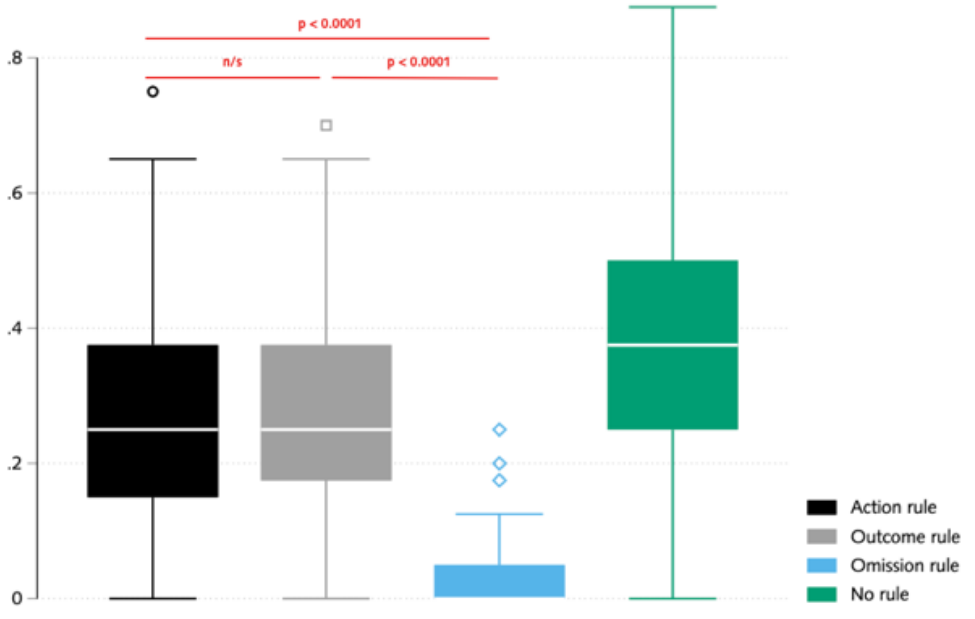


Figure S7—Levels of support within each group for the three possible rules, all rounds.

Anticipated support for the action rule

In addition to subjects' individual decisions to support the action rule, we measured subjects' expectations regarding which rule would be most supported by the other participants. Again, we hypothesized that the action rule would be supported significantly more than either the outcome rule or the omission rule.

The pattern of expectations was extremely similar to the pattern of actual decisions made (Figure S8). Anticipated support was not significantly different for the Action rule than for the Outcome rule ($z = -.573$, $p = 0.5666$) and both these rules were significantly more supported than the omission rule ($z = 4.604$, $p < 0.0001$).

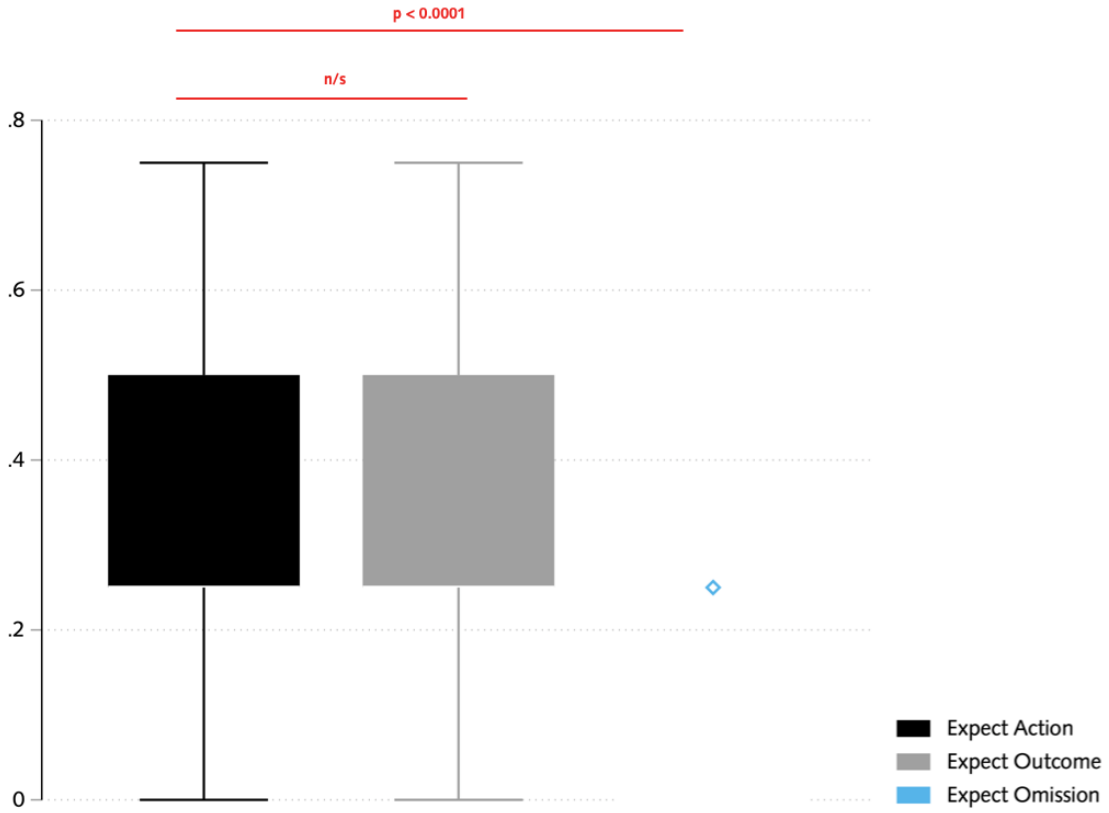


Figure S8—Average levels of expected support in each group for each punishment policy. (1 means that all four members of a group expected the same policy to be supported.)

Post experimental survey: Responses to acts versus omissions

In addition to our behavioral measure of punishment policies supported, we asked subjects how fair they believed it was to punish someone for a harmful act versus a harmful omission. We asked whether they would expect others to be angry at someone who committed a harmful act or a harmful omission; and we asked whether, if they were

themselves punished for a harmful act or omission, would they be likely to retaliate. For all three types of question, there was a significant difference between the response given for harmful acts versus harmful omissions, in the direction one would expect if there were an act–omission effect (Figure S9). Subjects thought it fairer to punish harmful acts, expected more anger at those who committed harmful acts, and were less likely to counter punish if they were punished for a harmful act. (Fairness of punishment: $z = 8.406$, $p < 0.0001$; Anger at transgressor: $z = 7.303$, $p < 0.0001$; Likely to counterpunish: $z = -3.493$, $p = 0.0005$. All Wilcoxon signrank tests at the individual level.)

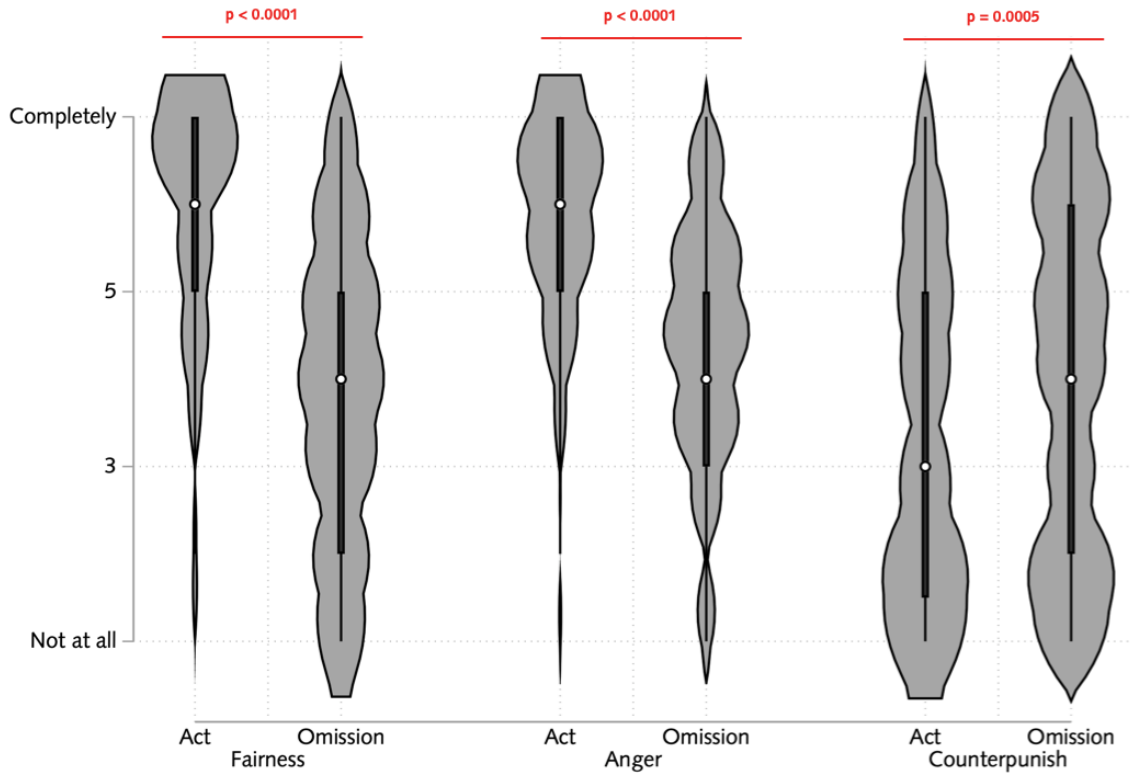


Figure S9—Violin plots of responses to questions asking: (i) does participant agree that it would be fair to be punished for having committed a harmful act/omission? (ii) does participant anticipate others would be angry at a participant who committed a harmful act/omission? (iii) would participant counterpunish someone who punished the participant for a harmful act/omission?

Discussion

In this study we employed a novel experimental paradigm to assess the degree to which subjects would enforce punishment of actors, omitters, or both, in a setting where there was a real cost associated with punishment and where both acts and omissions were simultaneously candidates to be punished. We found that punishing both acts and omissions was preferred roughly to the same degree as punishing acts alone. This result is still consistent with bias toward punishing acts, in that support for punishing omissions

alone was almost non-existent. Nonetheless, we had not anticipated so much enthusiasm for a punishment policy that treated the two types of harmful behavior as equivalent.

We were concerned that the results of this study may have been influenced by the fact that coordination was particularly difficult. Subjects had to make punishment decisions simultaneously, without knowing anything about the preferences of the agent with whom they would need to reach agreement. We thought it possible that the symmetric nature of the outcome rule made this a more salient coordination point than it would otherwise be and distorted the results. To remedy this, in Preliminary Study 2, we modified our initial design to make coordination more straightforward.

Preliminary Study 2

In **Preliminary Study 2**, we attempted to replicate the results of our first study, but in a setting where it is less difficult for the participants to successfully coordinate. Preliminary Study 1 required simultaneous coordination without communication, but in Preliminary Study 2 the punishment decision was made sequentially. The bystander proposed a punishment policy, which the victim could then endorse or reject.

Methods

This study used the same method as Preliminary Study 1, with the exception that the Button Affected Participant was asked not merely to indicate their preferred punishment policy, but whether they would endorse or accept each of the possible policies which the Unaffected Participant might propose first. If the Button Affected Participant endorsed the policy which was actually proposed, it would be enacted. See Figure S10.

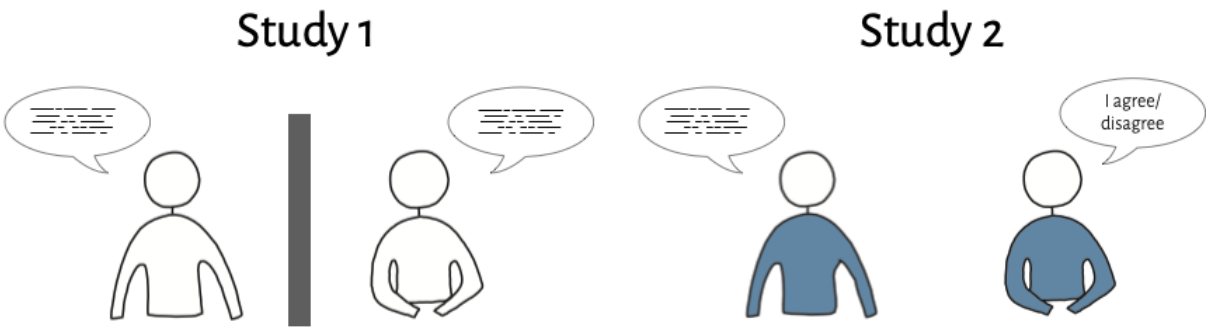


Figure S10. Difference between the procedures for determining punishment in Preliminary Studies 1 and 2. In Preliminary Study 1, each participant had to choose a punishment policy in ignorance of what their coordination partner was suggesting, and hope to achieve agreement. In Preliminary Study 2, the bystander made an initial proposal which the victim could choose to agree/disagree with.

Participants: 35 F, 41 M. Average age 21.6 years, sd = 2.5. Average earnings AUD30.67 (sd = 4.06).

Results

The average level of support for punishing acts versus omissions in each group was .56 versus .48 of group members (sd = .18, .17 respectively). Although this difference is not large in absolute terms, a within groups test rejects the null hypothesis that the probabilities of punishing acts and omissions are equal (Wilcoxon signrank test, $z = 2.592$, $p = 0.0095$).

The levels of support for each rule are summarized in Figures S11 and S12 below.

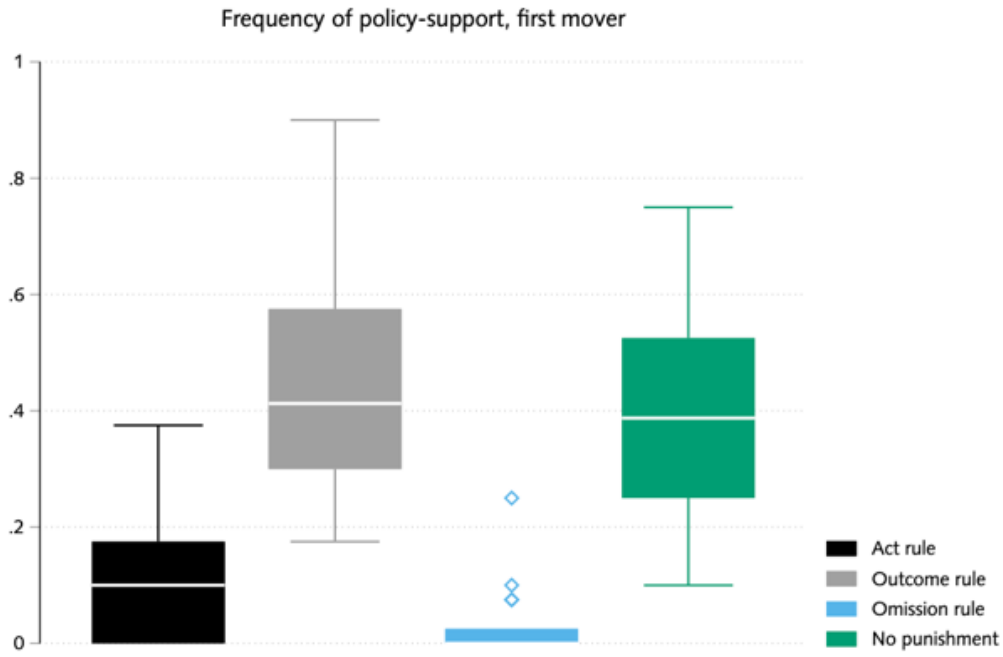


Figure S11—Frequency of support for various policies as first mover (rounds 5–9 only)

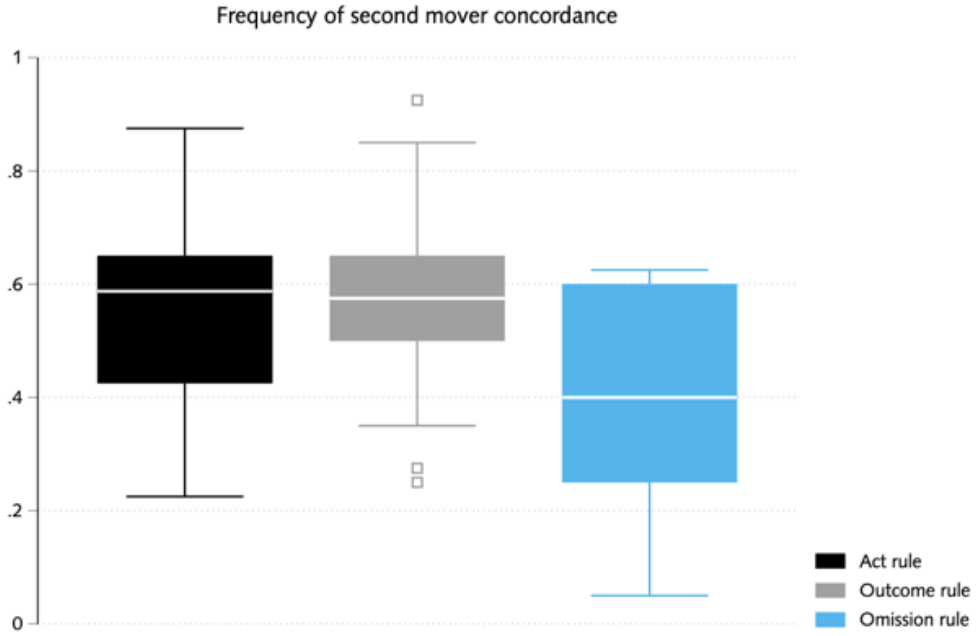


Figure S12—Frequency with which second mover supported first mover's suggested policy, by policy (rounds 5–9 only)

We also asked participants what policy they would support if they were able to make the decision individually, without relying on collective agreement (Figure S13). The results were very similar to the pattern observed in the first mover policy support.

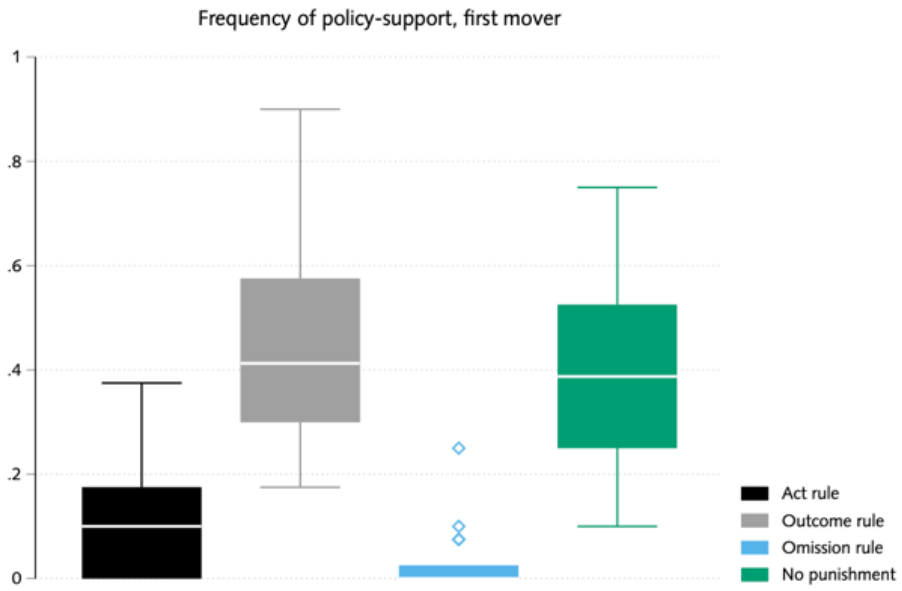


Figure S13—Frequency of support for policies, when making individual decision

We had two principal hypotheses for this study. The first was that, because we believe the act–omission distinction is influenced by the need to coordinate punishment, we predicted that the action rule would be supported more frequently in collective decisions than in individual decisions. More precisely, we predicted the probability that a group would collectively select the action rule to be enforced, conditional on selecting any punishment at all, would be greater than the probability that an individual in that group would select the action rule, conditional on selecting any punishment at all. This hypothesis was not supported. Only 7 out of 18 groups manifested the predicted asymmetry, with 6 having the opposite tendency (one sided binomial test, $p = 0.3872$).

Our second and third hypotheses together state that the frequency of collective support for the action rule would be greater than frequency of collective support for any other rule. Our key variables were the probabilities that a group would endorse each rule, conditional on their supporting any rule at all, (given all of their decisions, and given that each individual has an equal probability of being in the roles of Unaffected Participant and Button Affected Individual). Our hypotheses were:

> H2: Probability of enforcing the Action rule > Probability of enforcing the Outcome rule

H3: Probability of enforcing the Action rule > Probability of enforcing the Omission rule

While we found evidence that the Action rule is more likely to be selected than the Omission rule ($p = 0.004$), we did not find that the Action rule was more likely to be selected than the Outcome rule. Indeed, in 17 out of 18 groups, we found the opposite: the outcome rule was more likely to be supported than the Action rule ($p = 0.0002$, two-tailed probability that the two policies are equally popular). Thus, notwithstanding the modest sample size, the evidence offers some support for the notion that the most popular enforcement policy is the outcome rule, followed by the action rule, followed by the omission rule.

Discussion

In this study we found further evidence that the act–omission distinction does not lead to a clear preference to enforce punishment against actors only. We made it easier for participants to coordinate, reducing the need to rely on salient coordination points, and found that, if anything, there was even stronger support for punishing both actors and omitters than in Preliminary Study 1.

That said, these results are still consistent with an act–omission bias, given that there was continuing evidence that participants prefer to enforce the action rule rather than the omission rule.

Two further features of these preliminary studies gave rise to doubts about the external validity of our findings. First, in both studies, we used the strategy method to elicit punishment decisions: asking subjects to specify in advance how they would punish in response to every possible scenario. There is some evidence that "hot" punishment decisions are different from those made "cold" (Brandts and Charness 2011), so (notwithstanding that the evidence is mixed on this point, (Johnson and Mislin 2011; Fischbacher, Gächter, and Quercia 2012)), given the novelty of our design, we wanted to make sure that this was not responsible for our surprising results.

Second, the economic payoffs faced by our participants meant that the outcome rule could be enforced at no additional marginal cost, compared to enforcing the action rule. In effect, the outcome rule is a "punish one, get one free" option for those who wish to indulge in punishment. In normal life, the marginal cost of punishing an additional transgressor is almost always positive, and often increasing. For instance, detaining the last of a group of villains is usually more costly than detaining the first, because this is probably the outlaw who is best at evasion. Because of these additional concerns, we designed **Study 1** to remove both these constraints.

Pre-registered hypotheses

Below we give a complete account of all the preregistered hypotheses we tested across this project.

Preliminary Study 1

Table S8. Preregistered hypotheses and statistical test results for Preliminary Study 1. Results which pass our preregistered threshold significance level, including Holm–Bonferroni correction, are marked with an asterisk.

Hypothesis number	Null hypothesis	Z score (Wilcoxon ranksum)	P value
H2a	P[ALL,Action] = P[ALL,Outcome]	-0.098	0.9217
H2b	P[ALL,Action] = P[ALL, Omission]	4.753	<0.0001*
H4a	N[ALL,Action] = N[All,Outcome]	-0.573	0.5666
H4b	N[All,Action] = N[All,Omission]	4.604	<0.0001*

The same pattern of significance is obtained if we run these tests again, but including groups that were excluded because of failing our preregistered comprehension criterion.

Preregistered hypotheses 1 and 3 relate to other treatments that were not conducted, because our surprise at the results from the ALL treatment led us to abandon this design and commence our second preliminary study.

Preliminary Study 2

Table S9. Preregistered hypotheses and statistical test results for Preliminary Study 2. Results which are significant, given our preregistered intended alpha level, using Holm-Bonferroni correction, are marked with an asterisk.

Hypothesis number	Null hypothesis	Z score (Wilcoxon ranksum)	P value
H1	Pr(Second_action | Second_X) = Pr(Indvdl_action | Indvdl_X)	0.712	0.48
H2	Pr(First_action & Second_action) > Pr(First_outcome & Second_outcome)	–3.68	0.9998 (one-sided)
H3	Pr(First_action & Second_action) > Pr(First_omission & Second_omission)	2.868	0.004*
H4 (ancillary, not subject to Holm-Bonferroni adjustment)	Pr(First_action | First_X) > Pr(Indvdl_action | Indvdl_X)	20.71	0.038

The same pattern of significance is obtained if we run these tests again, but including groups that were excluded because of failing our preregistered comprehension criterion.

Study 1

Note that in our preregistration we registered hypotheses relating to the amount of punishment *accepted* rather than the amount of punishment *proposed*. We subsequently realized that the more revealing measure, given our theoretical interest, was amount of punishment proposed, and accordingly in the main text, we present analyses relating to this measure. (Amount of punishment accepted is of interest also, but calls for a more complex analysis, taking into account the amount of punishment proposed as further predictor.) To fulfil our preregistered plans for data analysis, we report all the originally preregistered tests on *accepted* punishment in Table S10. We also report, in Table S11, the same tests on amount of punishment *proposed*. The pattern of results is broadly similar: using group level tests to compare means at the group level, we have strong evidence to reject the null hypothesis corresponding to H3 only: that there is an act–omission effect. This corresponds to the main finding from our regression analysis in the paper.

Table S10. Preregistered hypotheses and statistical test results for Study 1, where the dependent variable is the amount of punishment accepted. For reasons noted in the text above, we no longer regard this as the appropriate choice of dependent variable, but are reporting these results for consistency with our preregistration. Because our main hypotheses are logically complex, the table breaks them down into their components. The main hypotheses are in ***bold****. After correcting for multiple hypothesis testing, none of the results achieve our preregistered significance level to control overall alpha for the main hypotheses at 0.05.*

Hypothesis number	Null hypothesis	Difference (SE)	95% CI	p(null)	P(null), including comprehension fails
H1a	A[red,2A] = A[green,2O] *Jointly responsible actions punished the same as jointly responsible omissions.*	1.24 (0.65)	–.06 to 2.54	.06	.07
H1b	A[green,2A] = A[red,2O] *Solely responsible actions punished the same as solely responsible omissions.*	0.42 (0.63)	–.85 to 1.69	.51	.65
H1	**H1a & H1b** *Jointly responsible punished the same; Solely responsible punished the same, regardless of act/omission.*			**p ≤ 0.06**	**p ≤ 0.07**
H2a(i)	A[green,2A] = A[green,2O] *Solely responsible omissions punished the same as jointly responsible omissions*	0.80 (0.64)	–.47 to 2.08	.21	.30
H2a(ii)	A[red,2A] = A[red,2O] *Solely responsible actions punished the same as jointly responsible actions*	0.86 (0.65)	–.43 to 2.15	.19	.22
H2b(i)	ΣA[green,2A] = ΣA[green,2O] *Aggregate expenditure on punishing jointly responsible omissions same as aggregate expenditure on punishing solely responsible omissions.*	–0.49 (0.92)	–2.34 to 1.36	.60	.37
H2b(ii)	ΣA[red,2A] = ΣA[red,2O] *Aggregate expenditure on*	3.39 (1.11)	1.17 to 5.62	.004	.002

	punishing jointly responsible actions same as aggregate expenditure on punishing solely responsible actions.				
H2	**(H2a(i) & H2a(ii)) or (H2b(i) & H2b(ii))**			**p ≥ 0.19**	**p ≥ 0.22**
H3	**A[red,2A] ≤ A[green,2O]** *Jointly responsible actions punished less than or the same as jointly responsible omissions*	**1.24 (0.65)**	**-0.06 to 2.54**	**0.03 (one-tailed)**	**0.04**

Table S11. Study 1, results of equivalent tests on proposed, rather than actual, punishment levels. Main hypotheses in bold.

Hypothesis number	Null hypothesis	Difference (SE)	95% CI	p(null)	P(null), including comprehension fails
H1a	P[red,2O] = P[green,2A]	0.30 (0.73)	–1.16 to 1.76	.68	.49
H1b	P[green,2O] = P[red,2A]	–1.35 (0.74)	–2.83 to 0.12	.07	.08
H1	**H1a & H1b**			**p ≤ 0.07**	**p ≤ 0.08**
H2a(i)	P[green,2A] = P[green,2O]	–0.61 (0.73)	–2.08 to 0.86	.41	.54
H2a(ii)	P[red,2A] = P[red,2O]	–0.45 (0.73)	–1.92 to 1.02	.54	.65
H2b(i)	ΣP[green,2A] = ΣP[green,2O]	1.57 (1.10)	–0.64 to 3.78	.16	.06
H2b(ii)	ΣP[red,2A] = ΣP[red,2O]	–3.98 (1.22)	–6.43 to –1.54	.002	.001
H2	**(H2a(i) & H2a(ii)) or (H2b(i) & H2b(ii))**			**p ≥ 0.41**	**p ≥ 0.54**
H3 (one-tailed version of H1b)	**P[red,2A] ≤ P[green,2O]**	**1.35 (0.74)**	**–0.12 to 2.83**	**.036**	**.040**

References

Brandts, Jordi, and Gary Charness. 2011. "The Strategy versus the Direct-Response Method: A First Survey of Experimental Comparisons." *Experimental Economics* 14 (3): 375–98. https://doi.org/10.1007/s10683-011-9272-x.

Chen, Daniel L., Martin Schonger, and Chris Wickens. 2016. "OTree—An Open-Source Platform for Laboratory, Online, and Field Experiments." *Journal of Behavioral and Experimental Finance* 9 (March): 88–97. https://doi.org/10.1016/j.jbef.2015.12.001.

Fischbacher, Urs, Simon Gächter, and Simone Quercia. 2012. "The Behavioral Validity of the Strategy Method in Public Good Experiments." *Journal of Economic Psychology* 33 (4): 897–913. https://doi.org/10.1016/j.joep.2012.04.002.

Johnson, Noel D., and Alexandra A. Mislin. 2011. "Trust Games: A Meta-Analysis." *Journal of Economic Psychology* 32 (5): 865–89. https://doi.org/10.1016/j.joep.2011.05.007.

Krupka, E L, and R A Weber. 2013. "Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?" *Journal of the European Economic Association* 11 (3): 495–524.

Selten, Reinhard. 1967. "Die Strategiemethode Zur Erforschung Des Eingeschränkt Rationalen Verhaltens Im Rahmen Eines Oligopolexperiments." In *Beiträge Zur Experimentellen Wirtschafts-Forschung.*, edited by H Sauermann, 136–68. Tubingen: Mohr.