Dato, Simon; Friehe, Tim

Punishment for intentions or outcomes: the role of gender and social norms

Social Choice and Welfare

Provided in Cooperation with:
Springer Nature

Suggested Citation: Dato, Simon; Friehe, Tim (2025) : Punishment for intentions or outcomes: the role of gender and social norms, Social Choice and Welfare, ISSN 1432-217X, Springer, Berlin, Heidelberg, Vol. 65, Iss. 4, pp. 853-882, https://doi.org/10.1007/s00355-025-01596-9

This Version is available at:
https://hdl.handle.net/10419/333357

Social Choice and Welfare (2025) 65:853–882
https://doi.org/10.1007/s00355-025-01596-9

ORIGINAL PAPER

Punishment for intentions or outcomes: the role of gender and social norms

Simon Dato[1] · **Tim Friehe**[2]

Received: 26 January 2023 / Accepted: 16 March 2025 / Published online: 11 April 2025

Abstract
Individuals often evaluate others' actions based on both their perceived intentions and their resulting outcomes, rewarding favorable actions and punishing unfavorable ones. This study aims to isolate the influence of these factors on punishment. We experimentally demonstrate that, when outcomes are held constant, second movers punish first movers who choose selfish actions more severely than those who select considerate ones. Conversely, when intentions are fixed, the severity of punishment does not significantly differ between fair and unfair outcomes on average. However, this average masks gender-specific variations. Men tend to prioritize punishing unkind intentions, while women are more sensitive to the perceived fairness of outcomes. Social norms help explain punishment choices and gender differences.

1 Introduction

Costly punishment plays a pivotal role in fostering cooperation and sustaining the well-being of societies, as demonstrated by extensive research (e.g., Gürerk et al. 2006; Balafoutas et al. 2014). Yet, the underlying motivations for individuals to engage in costly punishment remain a subject of ongoing inquiry (e.g., Colman 2006). Empirical evidence suggests that individuals are more likely to punish those who violate clear-cut social norms, particularly when the consequences of these violations are unambiguous (e.g., Fehr and Fischbacher 2004). However, when ambiguity in this respect emerges, such as in scenarios involving chance, individuals tend to base their punishment decisions also on perceived intentions rather than solely on outcomes (e.g., Falk and Fischbacher 2006).

While both outcomes and intentions can influence punishment choices, understanding their relative importance at the aggregate and individual levels is crucial

✉ Tim Friehe
tim.friehe@uni-marburg.de

Simon Dato
simon.dato@ebs.edu

1 EBS University for Business and Law, Rheingaustr. 1, 65375 Oestrich-Winkel, Germany

2 University of Marburg, Am Plan 2, 35037 Marburg, Germany

for predicting behavior and informing effective policy interventions. For instance, understanding employee preferences regarding procedural fairness or equitable outcomes can inform effective leadership decisions. More generally, the interplay between procedural and outcome justice is a critical factor in various economic and political contexts (e.g., Bolton et al. 2005; Cappelen et al. 2007). Existing research (e.g., Charness 2004; Falk et al. 2008) provides valuable insights, but further investigation is needed to understand the precise interplay between these factors and their individual-level heterogeneity.

To elucidate the relative importance of intentions and outcomes in shaping punishment, we conducted an experiment where first movers chose between two lotteries with identical outcomes but differing probabilities. The *Considerate* lottery offered a higher likelihood of an equal split, while the *Selfish* lottery favored an unequal split, benefiting the first mover. Second movers, aware of the first mover's choice and the resulting outcome, could punish their first mover. This experimental design allows us to isolate the effects of intentions and outcomes on punishment decisions. Fixing the outcome and comparing punishment across different lottery choices, we demonstrate how unkind instead of kind intentions influence punishment. Likewise, holding constant the first-mover's lottery choice (i.e., her intention) and comparing punishment across outcomes, we isolate how the unequal compared to the equal outcome affects the second-mover's punishment choice.

Our findings reveal significantly more punishment when first movers select the selfish lottery than when they select the considerate one, indicating a strong aversion to unkind intentions. In contrast, when intentions are held constant, punishment severity does not vary when the outcome changes. This does not imply that outcomes are irrelevant, as the insignificance can also stem from counterbalancing effects such as income effects.

Given the extensive literature on gender-specific social preferences (e.g., Eckel and Grossman 2008; Croson and Gneezy 2009; Niederle 2016), our study investigates the potential impact of gender on the relative valuation of outcomes and intentions when making punishment decisions. Understanding gender differences in fairness perceptions is crucial for addressing societal and organizational inequalities. As workplaces and institutions strive for gender balance, it is imperative to recognize that women may have distinct expectations regarding fairness. Recognizing these differences allows for developing more suitable compensation and promotion schemes. Our findings reveal that women and men respond differently to unkind intentions and unequal outcomes. Women are more likely to increase their punishment levels in response to unequal outcomes, suggesting a greater concern for fairness. Conversely, men are more inclined to punish selfish intentions, indicating a heightened sensitivity to unkind intentions.

To investigate the underlying drivers of punishment behavior, we implemented a cooling-off period before the second movers' decisions. This manipulation aimed to reduce emotional involvement, following established research (e.g., Cardella and Chiu 2012; Neo et al. 2013). Given the connection between punishment and emotional expression (e.g., Xiao and Houser 2005), we hypothesized that a cooling-off period would lower punishment. Additionally, considering the literature on gender differences in emotional involvement (e.g., Croson and Gneezy 2009; Fujita et al. 1991), we anticipated a stronger impact of the cooling-off period on women's punish-

ment decisions. However, our results did not reveal a significant treatment effect on punishment for either men or women.

To delve deeper into how social norms shape punishment decisions, we employed the methodology outlined by Krupka and Weber (2013). A growing body of research suggests that a preference for choosing socially appropriate actions can significantly influence economic behavior. In our context, second movers may be guided by perceived social norms regarding punishment, leading them to choose socially acceptable levels of punishment.

By eliciting individuals' perceptions of the social appropriateness of punishment, we demonstrate that a preference for norm compliance plays a significant role in shaping their choices. Furthermore, gender-specific social norms can partially explain the observed gender differences in punishment behavior. This suggests that understanding the interplay between individual preferences and societal expectations is essential for fully comprehending the dynamics of social punishment.

Our research contributes to the existing literature in three key ways. First, we provide a comprehensive analysis of the relative importance of intentions and outcomes in shaping punishment decisions, highlighting the gender-specific nature of these preferences. While previous studies (e.g., Bolton et al. 2005; Falk et al. 2008; Friehe and Utikal 2018) have emphasized the distinction between outcome-based and intention-based punishment, our research offers novel insights into how these factors differ across genders. Second, we demonstrate a strong alignment between the punishment behavior observed in our data and the prevailing social norms. Our findings contribute to the growing body of research on the relationship between social norms and decision-making, particularly in the context of punishment (Barr et al. 2018; Chang et al. 2019). Third, we address the ongoing debate raised by Fehr et al. (2018) regarding the existence and relevance of social norms in shaping punishment choices. Our results provide compelling evidence supporting their influence on decision-making processes.

The paper's structure is as follows. In Sect. 2, we discuss the related literature. We explain the experimental design and procedures in Sect. 3. Section 4 presents our main hypotheses. Section 5 reports our empirical findings regarding actual punishment choices and social norms on punishment choices. Section 6 concludes.

2 Related literature

This paper examines the relative importance of intentions and outcomes in shaping individual behavior. Prior research has utilized two primary approaches to isolate intention effects: (i) comparing participant responses to the choices of others with their reactions to actions imposed exogenously (e.g., Blount 1995; Charness 2004; Cox 2004; Falk et al. 2008), and (ii) analyzing choices made at specific decision nodes within a game based on how these nodes were reached (e.g., Falk et al. 2003; McCabe et al. 2003).

Blount (1995) analyzes ultimatum games and compares second-movers' minimum acceptable offers. She varies whether the offer is implemented by a self-interested party, a third party, or by chance. She finds that second-movers are more willing

to accept a low offer if implemented by chance than by a self-interested party. In Cox (2004), decision-making in a standard trust game is compared to choices when allocations are exogenously imposed to match interim outcomes of the standard trust game. In other words, subjects in this latter treatment do not *respond* to a first mover. Similarly, Charness (2004) studies a gift-exchange game and analyzes second-mover behavior when the employer selects wages compared to an external process. In the moonlighting-game setup of Falk et al. (2008), first-mover choices were either chosen deliberately by a subject in the *Intentions* treatment or randomly drawn in the *No Intentions* treatment. In these contributions, the second-mover reaction is stronger when the first-mover choice is intentional. Despite this commonality, the different papers feature design differences that may be important. For example, in Charness (2004), efficiency motives possibly interacted with pure reciprocity concerns in a setting featuring repetition. In contrast, the game in Cox (2004) was played once, and the second mover decided about an efficiency-neutral transfer. Compared to the other two contributions, Falk et al. (2008) allows for positive and negative reciprocity.

Our setup diverges from those employed in previous research, as we do not implement separate treatments to isolate intention effects. Our design lacks an intention-free condition. Instead, interim outcomes are jointly determined by the intentional choices of a first-mover and the realization of a random variable. Consequently, a specific outcome can arise from a variety of first-mover decisions. We capture the *intention effect* by examining how changes in first-mover behavior influence subsequent responses, holding the outcome constant. This approach contrasts with the predominant methodology in existing literature, which primarily compares reactions to intentionally implemented and randomly drawn outcomes.

The distinction between our approach and previous contributions is important. For example, the previous literature in neuroeconomics has shown that participants' emotional response to a human act is very different from the response to an act by the computer (e.g., Rilling and Sanfey 2011). For example, van't Wout et al. (2006) find that unfair offers in an ultimatum game triggered higher skin conductance activity and rejection rates only if the offer came from a human proposer. Our intention-effect identification maintains this emotional influence by comparing responses to kind and unkind first-mover behavior, whereas comparisons of responses to first-mover and random choices do not. This seems particularly relevant in light of studies identifying the role of punishment for emotional expression (e.g., Xiao and Houser 2005).

Furthermore, our approach allows us to assess the *outcome effect* within the context of authentic first-mover decision-making rather than relying on comparisons between reactions to arbitrarily drawn outcomes. In this regard, our study resembles that of Charness and Levine (2007), where principals initially select between a high and a low wage. Crucially, both wage choices can ultimately result in an intermediate wage for the agent due to the realization of a random event. This feature enables investigating how a single outcome can be achieved through different first-mover intentions. Charness and Levine (2007) demonstrate the significance of both outcomes and intentions in shaping reciprocal behavior, with intentions exerting a stronger influence. However, their setup differs from ours in a crucial aspect: the lotteries associated with high and low wages involve distinct outcome distributions. In contrast, our design employs lotteries that differ only in their probability distributions while maintaining identical

outcome sets. This distinction may render the lottery choice less salient and potentially influence how subjects interpret the first-mover's intentions.

We find that outcomes are relatively more important for women, whereas the reverse is true for men. We are the first to separate the relative importance of intentions and outcomes by gender, even though a large strand of the literature analyzes potential gender differences in social preferences. For example, Croson and Buchan (1999) study gender differences in trust games, finding that sender behavior is similar across genders while women returned a higher proportion of their wealth. Buchan et al. (2008) find that men trust more while women are more trustworthy. In contrast, Chaudhuri and Gangadharan (2007) find no differences in reciprocal behavior. Using the ultimatum game, Eckel and Grossman (2001) find that women are more likely to accept lower offers than men. Overall, the literature presents quite mixed results (e.g., Croson and Gneezy 2009), meaning that more evidence is needed. Whereas previous contributions use experimental paradigms in which an action's intentions and consequences were inextricably linked (e.g., the ultimatum game), our experiment disentangles intentions and outcomes. This allows us to cleanly identify gender differences in the importance of (i) outcomes and (ii) intentions on punishment.

Our results also contribute to the recent and growing literature documenting the explanatory power of social norms for observed behavior. We build on Krupka and Weber (2013) and find that even gender differences in punishment decisions can be partly explained by reference to gender-specific social norms. Similar to Barr et al. (2018) and Chang et al. (2019), we can thus show that perceptions of social norms are identity-specific.

Our treatment variation includes a cooling-off period. This follows contributions such as Grimm and Mengel (2011). Using an ultimatum game, they find that a delay of around 10 min after the presentation of the offer and before the final acceptance choice causes a significant increase in the acceptance rate of low offers. Whereas most studies (e.g., Cardella and Chiu 2012; Neo et al. 2013) also consider a relatively short delay, Oechssler et al. (2015) study how a 24-hour delay influences ultimatum-game play, distinguishing a treatment in which subjects are paid in cash from one in which they are compensated with lottery tickets. They find that the cooling-off period influences rejection choices only when subjects receive lottery tickets. The fewer rejections that Neo et al. (2013), for example, find in their delay-treatment of the ultimatum game align with the idea that immediate decisions show the participants' intuitive responses to the inequity. In contrast, delayed decisions result after more careful deliberation about monetary payoff consequences. Similarly, our cooling-off period was expected to reduce punishments. However, deliberation may also increase punishment. The data in Philippsen et al. (2024) is consistent with the idea that the participants' intuitive response is a selfish payoff maximization and that a preference for costly punishment emerges only with time to deliberate.

3 Design

The experiment consisted of two primary parts. In Part 1, Player A selected one of two lotteries, each offering a different probability of an equal or unequal outcome.

Player B, aware of Player A's choice and the resulting payoff allocation, then assigned punishment points. In Part 2, following the methodology of Krupka and Weber (2013), we elicited participants' perceptions of social norms related to the game. Part 3 involved a questionnaire, including an incentivized social value orientation test and a survey on participants' justice attitudes. Additionally, we employed the experimental task developed by Kimbrough and Vostroknutov (2018) to assess participants' adherence to social norms.

At the outset, participants were informed about the study's structure, including the existence of Part 1 and the subsequent payoff-independent parts. In line with Dato and Nieken (2014), we collected demographic information from our subjects before Part 1 to enable gender-specific matching in Part 2. Our experiment featured two treatments: DELAY and IMMEDIATE. We will first describe treatment IMMEDIATE and then outline the key differences in treatment DELAY.

3.1 Part 1: First-mover's lottery choice and second-mover's punishment choice

Part 1 comprised two stages. In Stage 1, Player A selected either the *Selfish* (abbreviated S) or the *Considerate* (abbreviated C) lottery. Lottery *Selfish* led to the *unequal* payoff allocation (abbreviated U) $(\pi_U^A, \pi_U^B) = (1350, 150)$ with probability 80% and the *equal* payoff allocation (abbreviated E) $(\pi_E^A, \pi_E^B) = (750, 750)$ with probability 20%. Lottery *Considerate* reversed the probabilities (i.e., it yielded the unequal payoff allocation with probability 20%). Player A could not dictate a division of the endowment amounting to 1500 points but skew the probability distribution towards the unequal or the equal payoff allocation.

Player A's expected payoff exceeded Player B's in both lotteries. With *Selfish*, Player A (B) expects 1230 (270). With *Considerate*, Player A (B) expects 870 (630). Regarding the inequity aversion model of Trautmann (2009), the choice of *Selfish* substantially increases Player B's disadvantageous inequity.

At the end of Stage 1, Player B was informed about Player A's lottery choice and the lottery's outcome. Thus, a randomly matched pair of Players A and B had common knowledge about which one out of four possible scenarios is relevant to them:

- Scenario SE: Player A's choice of *Selfish* combined with the draw of the *equal* payoff allocation,
- Scenario SU: *Selfish* combined with *unequal* payoffs,
- Scenario CE: *Considerate* combined with the draw of the *equal* payoff allocation, or
- Scenario CU: *Considerate* combined with *unequal* payoffs.

Knowing the relevant scenario, Player B could deduct p points from Player A's account in Stage 2 at a cost $p/4$.[1] Subjects could choose a punishment level p, where

$$p \in \{0, 60, 120, 180, 240, 300\}.$$

The maximum punishment level is sufficiently high to allow B to spend a sizable share of her interim payoff on punishment. It is, however, also sufficiently low to allow a clear role of punishment depending upon the scenario: punishment increases advantageous payoff inequality in SE and CE, and decreases disadvantageous payoff inequality in SU and CU. This allows deriving clear-cut predictions regarding the outcome effect based on the model's primitives introduced by Charness and Rabin (2002) in Sect. 4. To enable emotional involvement (one of the hypothesized channels for gender differences), we purposefully implemented a direct-response format and relied on between-subject comparisons (Brandts and Charness 2011). The final payoffs amount to

$$\Pi_i^A(p) = \pi_i^A - p \quad \text{and} \quad \Pi_i^B(p) = \pi_i^B - \frac{p}{4},$$

where $i \in \{E, U\}$ depicts the outcome drawn in Stage 1. Before making their decisions in Part 1, participants provided incentivized belief statements. Player A indicated the expected punishment level from their respective Player B across the four scenarios. Conversely, Player B stated their anticipated lottery choice. Participants earned 200 points for each accurately predicted choice.

When collecting choice and belief data from the same subject, the order of elicitation becomes a crucial consideration. Schlag et al. (2015) provide a comprehensive review of relevant research, examining various experimental paradigms. This review concludes that the impact of belief elicitation on subsequent decisions remains uncertain, both in terms of its presence and direction. Similarly, Alempaki et al. (2022), where participants ranked outcomes and considered beliefs before making choices, also found inconclusive evidence regarding the influence of prior belief elicitation.

In our setup, how an incentivized belief elicitation influences subsequent decisions is also unclear. To illustrate that the effect on punishment choices is ambiguous, consider the case where Player B correctly guessed Player A's lottery choice. As a result, B receives an additional payoff, which can have two implications. First, Player B may want to choose higher punishment due to the reduced marginal utility of income. However, second, the additional income might reduce negative emotions (if A chose the selfish lottery) or intensify positive emotions (if A chose the considerate lottery), making less punishment likely.

[1] We implemented a cost-effectiveness ratio of 1:4, which is similar to the 1:3 ratio used in Fehr and Fischbacher (2004) and Leibbrandt and Lopez-Perez (2012). Nikiforakis and Normann (2008) tested ratios 1:1, 1:2, 1:3, and 1:4 in a comparative-statics exercise, finding that ratios 1:3 and 1:4 perform similarly in numerous regards. Other papers use an even higher punishment effectiveness. For example, Bartling et al. (2014) used a 1:5 ratio. On a different note, we follow the literature by using the term punishment for Player B's action in all scenarios. Note that Player B's payoff change in scenarios involving *Considerate* as Player A's lottery choice presumably served distributional preferences and not a truly *punitive* motivation. In a stricter understanding, punishment is a hardship imposed on someone for a wrong they have (or are believed to have) committed (e.g., Bagaric 2001).

Importantly, as we pay particular attention to gender differences, note that there is no gender gap in the average payoff from the belief elicitation ($p = 0.226$, Fisher's exact (FE)). Even though these arguments and the findings from the related literature do not hint at a systematic effect of incentivized beliefs, we can, of course, not definitively rule out that punishment behavior was affected by the incentivized belief elicitation procedure.

3.2 Part 2: Elicitation of social norms

Following Barr et al. (2018), d'Adda et al. (2016), and Erkut et al. (2015), we elicited social norms from our participants regarding the game they played in Part 1.[2] Participants were asked to give 26 social appropriateness ratings: six punishment levels for the four scenarios possible plus the two lottery choices. The order of social appropriateness ratings was randomized at the subject level. We employed the six-point scale from Chang et al. (2019) comprising: "very socially appropriate" (later assigned a score of 5 in our empirical work), "socially appropriate" (4), "somewhat socially appropriate" (3), "somewhat socially inappropriate" (2), "socially inappropriate" (1), and "very socially inappropriate" (0).

The evaluation of choices was incentivized. One of the 26 choices was randomly selected, and each participant's evaluation of that choice was compared to that of another experimental subject (Barr et al. 2018; Erkut et al. 2015). If a participant's evaluation matched the other subject's rating, this participant earned 1200 points; otherwise, this participant earned nothing. Stipulating payoffs like this means that subjects play a coordination game where participants are incentivized to state the normative evaluation of their match. According to Krupka and Weber (2013), this scheme incentivizes participants to reveal their perception of what is commonly regarded as socially appropriate or inappropriate behavior in the context at hand instead of eliciting their private evaluation.

We informed subjects about the gender of their randomly matched subject before they made their ratings (producing observations from single-gender and mixed-gender pairs) to accommodate the possibility of *commonly known* gender-specific social norms. We can use this data to assess whether participants condition their rating on the matched player's gender. Importantly, gender differences in social norms that are not commonly known will not produce such an adjustment in response to the revelation of the randomly matched subject's gender.

3.3 Part 3: Questionnaire

To assess rule-following behavior, we employed the task introduced by Kimbrough and Vostroknutov (2018), which is considered a reliable measure of an individual's propensity to adhere to social norms (see, among others, Kimbrough and Vostroknutov 2016, 2018; Gross and Dreu 2021). Participants dragged and dropped 50 balls into one of two buckets: yellow or blue. Instructions clearly stated that the rule was to

[2] We elicited social appropriateness ratings in Part 2 that describe *injunctive* social norms. In contrast, the beliefs elicited in Part 1 concern *descriptive* social norms.

deposit balls in the blue bucket, with rewards of 6 points for each ball in the yellow bucket and 3 points for each ball in the blue bucket. Payoffs were directly determined by the number of balls placed in each bucket.

Following the rule-following task, participants completed a version of the Social Value Orientation slider measure (Murphy et al. 2011). This measure provides valuable insights into participants' preferences for balancing self-interest with the interests of others, a crucial factor likely influencing both Player A's lottery choices and Player B's punishment decisions.

Next, we assessed participants' justice sensitivity using the short scale developed by Baumert et al. (2014). This scale incorporates perspectives from both victims and offenders, offering a comprehensive understanding of individual justice concerns. As Schmitt et al. (2010) explain, justice sensitivity significantly impacts perceptions of and reactions to various forms of injustice. Therefore, this measure can provide valuable insights into Player B's responses to the presented scenarios and may also help explain Player A's selection of the Selfish lottery.

Finally, participants provided basic demographic information, including age, number of siblings, and course of study.

3.4 Treatment DELAY

Treatment DELAY differs from treatment IMMEDIATE only in that a set of questions, requiring about 10 min to answer, preceded Player B's punishment decision in Part 1's Stage 2. The questions stemmed from questionnaires used in other experiments and were unrelated to the two stages of Part 1. For example, we elicited participants' risk attitudes, CRT scores, and how often participants have previously participated in experiments. These questions were not incentivized. The delay introduced a "cooling off" period, which was shown to be consequential in some other experimental settings (e.g., Cardella and Chiu 2012).[3]

3.5 Procedures

The experiment was conducted at the BonnEconLab in December 2019, using the online recruitment hroot (Bock et al. 2014) and the experimental software z-Tree (Fischbacher 2007). Our sessions lasted about 1.5 hours. 380 subjects participated in our experiment. To enable different possibilities of matching genders in Part 2, we aimed at a gender balance at the session level. Overall, 191 males and 189 females took part. Parts 1–3 were payoff relevant. The average earnings amounted to 20.45 Euro, using an exchange rate of 1 point to 1 Euro Cent.

[3] Note that this procedure does not grant full control regarding emotions. A more direct way of addressing the influence of emotions is by asking participants questions. For example, Khadjavi (2015) and Reuben and Winden (2010) directly ask about 16 emotions using a seven-item survey scale. However, timing the elicitation so that it appears before the punishment decisions runs the risk of inducing an experimenter-demand effect, and placing it after the punishment choices allows for the possibility that emotional intensity has been reduced because of the punishment.

4 Behavioral predictions

We follow Charness and Rabin (2002) to derive behavioral predictions, allowing outcome- and intentions-based social preferences. Assume that Player B's utility function in scenario ji, with $j \in \{C, S\}$ and $i \in \{E, U\}$, can be stated as

$$U_{ji}(p) = u(\Pi_i^B) + \rho \min\{\Pi_i^A - \Pi_i^B, 0\} + \sigma \max\{\Pi_i^A - \Pi_i^B, 0\} + \theta q_{\text{Selfish}} \left(\Pi_i^B - \Pi_i^A\right), \tag{1}$$

where $q_{\text{Selfish}} = 1$ if $j = S$ (i.e., if Player A chose *Selfish*) and zero otherwise. Next to the utility from the material payoff $u(\Pi_i^B)$, where $u'(\cdot) > 0 \geq u''(\cdot)$, Player B's utility is affected by the difference between the own and Player A's material payoff. The attitude towards (dis)advantageous inequity is captured by ρ (σ). It is natural and in line with Charness and Rabin (2002) to assume $\rho \geq \sigma$, so that B's preference for gains relative to A is not stronger when being ahead than when being behind. As explained by Charness and Rabin (2002), *competitive preferences* arise with $\sigma \leq \rho \leq 0$, *inequity aversion* by assuming $\sigma < 0 < \rho < 1$, and *social-welfare preferences* with $0 < \sigma \leq \rho \leq 1$. Finally, intentions-based reciprocity is incorporated via $\theta \geq 0$: if A chose the *Selfish* lottery, B is more competitive towards A.

Since $\Pi_E^B \geq \Pi_E^A$ ($\Pi_U^B < \Pi_U^A$) holds for all punishment levels, only (dis)advantageous inequity is relevant in (un)equal-payoff scenarios. Inserting $\Pi_i^A = \pi_i^A - p$ and $\Pi_i^B = \pi_i^B - \frac{p}{4}$ and considering how U_{ji} changes with p, we obtain the following marginal utilities for scenarios with equal (jE) and unequal (jU) payoffs:

$$U_{jE}^{'}(p) = \frac{\theta q_{\text{Selfish}} - 3\rho - u'(\Pi_E^B)}{4} \quad \text{and} \quad U_{jU}^{'}(p) = \frac{\theta q_{\text{Selfish}} - 3\sigma - u'(\Pi_U^B)}{4}.$$

To understand how the outcome influences the desirability of punishment, we compare the marginal utilities across scenarios with different outcomes:

$$U_{jU}^{'}(p) - U_{jE}^{'}(p) = \frac{1}{4}\left[u'(\Pi_E^B) - u'(\Pi_U^B)\right] + \frac{3}{4}(\rho - \sigma). \tag{2}$$

The outcome effect can be separated into an *inequity effect* and an *income effect*, captured by the first term. Given that (i) Player B's payoff is strictly higher under the equal outcome and (ii) the utility function, $u(\cdot)$, exhibits weak concavity, the first term of the difference in marginal punishment incentives is weakly negative. Higher payoffs imply a lower marginal utility of income and, thus, a lower opportunity cost of investing in punishment. Via the income effect, punishment becomes less attractive after an unequal payoff draw.

The second term in (2), the inequity effect, is determined by the difference between ρ and σ. After the draw of (un)equal payoffs, B is (dis)advantaged in payoff terms, and the optimal punishment depends on ρ (σ). The lower σ (ρ) is, the more competitive or less altruistic is B towards A. With $\rho \geq \sigma$, the inequity effect renders punishment more attractive after the draw of unequal payoffs.

The inequity effect dominates the income effect when the utility function is not too concave within the relevant payoff range. Experimental evidence supports this dominance. In the study by Fehr and Fischbacher (2004) on the dictator game with second-party punishment, recipients experiencing disadvantageous inequity can mitigate this inequity through punishment. Crucially, among these disadvantaged recipients, those with lower initial allotments face greater inequity while simultaneously incurring higher punishment costs due to the concavity of their income utility function. Figure 5 in Fehr and Fischbacher (2004) demonstrates that recipients who received less from their dictators (i.e., poorer recipients) exhibited higher levels of punishment.

Regarding the intention effect, we obtain the following marginal utilities:

$$U_{SE}^{'}(p) = \frac{\theta - 3\rho - u'(\Pi_E^B)}{4}, \text{ and } U_{CE}^{'}(p) = \frac{-3\rho - u'(\Pi_E^B)}{4}$$
$$U_{SU}^{'}(p) = \frac{\theta - 3\sigma - u'(\Pi_U^B)}{4}, \text{ and } U_{CU}^{'}(p) = \frac{-3\sigma - u'(\Pi_U^B)}{4}.$$

To understand how Player A's intentions influence the desirability of punishment, we compare the marginal utilities of scenarios with identical outcomes but different lottery choices:

$$U_{Si}^{'}(p) - U_{Ci}^{'}(p) = \frac{3}{4}\theta \geq 0, \quad i \in \{E, U\}. \tag{3}$$

The positive impact of punishment on Player B's utility is larger after a selfish choice by Player A than a considerate one. Accordingly, we expect more punishment after selfish than after considerate choices.

Hypothesis 1 *(a) The outcome effect will be positive, i.e., punishment will be higher after a draw of the unequal instead of the equal outcome for a given lottery choice. (b) The intention effect will be positive: punishment will be higher for a given outcome after A's choice of the selfish instead of the considerate lottery.*

Next, we explore gender differences and start with the outcome effect. Results presented in Andreoni and Vesterlund (2001), Engel (2011), and Niederle (2016) suggest that women seem more concerned than men with equalizing payoffs in lab experiments. However, Eckel and Grossman (2001) show that women—in the role of receivers in an ultimatum game—are more likely than men to accept unequal offers with disadvantageous inequity. Bellemare et al. (2008) measure σ and ρ at the individual level in ultimatum games and do not find substantial gender differences. In summary, the evidence on gender differences regarding the parameters ρ and σ and, hence, the inequity effect is mixed. Regarding the income effect, women are generally more risk-averse than men (Croson and Gneezy 2009). Correspondingly, a more concave utility function for money could more strongly diminish the attractiveness of punishment following an unequal outcome for women. However, due to the opposing influences, we refrain from hypothesizing gender differences in the outcome effect.

Regarding a gender-specific intention effect, we focus on heterogeneity in the parameter θ and draw upon the work of Del Giudice et al. (2012). They found higher

levels of rule-consciousness for men in a large US sample, which suggests that men are more inclined to adhere to the principle of punishing unkind intentions. This aligns with the findings of Eckel and Grossman (1996) and the observation by Croson and Gneezy (2009), which suggest that men are (i) less sensitive to the situational context when making decisions and (ii) more likely to make principled decisions. Accordingly, we expect men to react more strongly than women to unkind intentions.

Hypothesis 2 *The intention effect is stronger for men than for women: for a given payoff allocation, men assign greater incremental punishment than women for Player A's choice of the selfish instead of the considerate lottery.*

Our experiment includes treatments DELAY and IMMEDIATE. The former is meant to lower affect intensity. Punishment is one way to express negative emotions (e.g., Dickinson and Masclet 2015; Hopfensitz and Reuben 2009; Joffily et al. 2014; Masclet and Villeval 2008; Xiao and Houser 2005). A cooling-off period can mitigate negative emotions and lead to more moderate punishment reactions (e.g., Dickinson and Masclet 2015; Grimm and Mengel 2011; Oechssler et al. 2015). Accordingly, we expect less punishment in DELAY. Regarding gender, an extensive body of literature reports that women tend to be more emotionally involved and expressive than men (e.g., Briton and Hall 1995; Simon and Nath 2004), which supports widespread stereotypes regarding the relationship of gender and emotions (e.g., Robinson and Johnson 1997; Timmers et al. 2003). The gender-specific emotional involvement is particularly relevant for our results in IMMEDIATE. Accordingly, we expect that the punishment-lowering effect of DELAY is more pronounced for women than for men.

Hypothesis 3 *(a) For both genders, punishment is lower in treatment DELAY than in treatment IMMEDIATE. (b) The dampening effect of treatment DELAY is more pronounced for women.*

Several contributions show that a preference for aligning choices with social norms can explain much of economic behavior (e.g., Krupka and Weber 2013; Kimbrough and Vostroknutov 2016; Chang et al. 2019). The preferences presented in Krupka and Weber (2013) can be stated as follows:

$$U_{ji}(p) = u\left(\Pi_i^B(p)\right) + \gamma N_{ji}(p), \tag{4}$$

where $\gamma \geq 0$, $N_{ji}(p) > (<) 0$ for (in)appropriate punishment levels in scenario ji, and N_{ji} increases with the social appropriateness of punishment p. Individuals maximizing the utility in equation (4) may sacrifice material payoff for a higher payoff from norm compliance. The parameter γ scales the relative importance of norm compliance. The rule-following task of Kimbrough and Vostroknutov (2018) included in our questionnaire attempts to measure this parameter. The literature has shown that cooperative behavior is a social norm (e.g., Fehr and Fischbacher 2004) and that the threat of punishment supports cooperation (e.g., Fehr and Gächter 2000). However, whether or not punishing is socially appropriate has not been tested.

Beyond examining punishment norms, we investigate the social norm of Player A's lottery choice and analyze the interplay between lottery choice and punishment norms, explicitly examining whether Player A's norm violations increase the perceived appropriateness of Player B's punishment. In other words, we test whether the motivation to punish norm violations is independent of, or picked up by, the social norm of punishment.

Recent norms research highlights the potential for both social norms (represented by N_{ji}) and the weight placed on norm compliance (represented by γ) to vary across individuals (Barr et al. 2018; Chang et al. 2019). Specifically, a gender-specific perception of punishment norms may emerge. For instance, if men prioritize principled choices more strongly, they may perceive incremental punishment for unkind intentions as more socially appropriate than women. Furthermore, research suggests that women are more likely to view anger as a socially acceptable emotion (Simon and Nath 2004). Consequently, they may perceive the punishment of actions that elicit anger as more socially appropriate than men.

Hypothesis 4 *Incorporating social appropriateness ratings of punishment and lottery choices helps explain punishment choices and gender differences in punishment.*

5 Results

In Sect. 5.1, we analyze lottery choices by Players A before we turn to our main interest, punishment levels as a function of intentions, outcomes, and gender. Section 5.2 analyzes social norms (elicited in Part 2 of the experiment) as a driver of actual punishment choices.

5.1 Lottery choice and punishment (part 1)

5.1.1 First-mover's lottery choice

Players A chose *Selfish* in 72.6% of all cases. The share of *Selfish* is not treatment-dependent (73.7% IMMEDIATE vs. 71.6% DELAY; $p = 0.871$, FE). Moreover, we find that the share of *Selfish* is not gender-dependent (75.5% men vs. 69.6% women; $p = 0.417$, FE).[4] These results are confirmed by probit regressions that yield the additional insight that Players A with a higher anticipated incremental punishment

[4] Different aspects influence decision-making in Stage 1. These aspects may influence the decision-making of genders in different ways. This can explain the absence of a gender effect regarding lottery choice. For example, the choice of *Considerate* may be considered more *prosocial*. The evidence in Engel (2011) suggests that women are more prosocial and, thus, should be more likely to choose *Considerate*. On a different note, choosing *Considerate* minimizes the expected inefficiency from punishment cost, which is the likelier anticipated by subjects more strategically sophisticated. Fehr et al. (2006) find that women are relatively more concerned about equity than efficiency and Cubel and Sanchez-Pages (2022) present evidence suggesting that men are weakly more strategically sophisticated. This dimension, thus, makes us expect that men are more likely to choose *Considerate*.

after a choice of *Selfish* were less likely to select the *Selfish* lottery (see our Appendix A for detailed regression results).[5] This shows that the punishment stage was important for first-stage decisions.

5.1.2 Second-mover's punishment

We do not find significant differences in punishment levels between the IMMEDIATE and DELAY treatments, neither overall nor when considering gender-specific effects.[6] This finding suggests that emotions were not a primary driver of punishment in our study. Men and women do not respond differently to the treatment (as indicated by the insignificant interaction of the treatment and gender dummy variables in Appendix B).[7] Therefore, we reject Hypothesis 3 and proceed by pooling the data from both treatments in our subsequent analyses.

Our main hypothesis concerns how unkind intentions and unequal outcomes causally influence punishment levels. To analyze these effects, we split our punishment data according to the possible combinations of intentions and outcomes. As Players A predominantly chose *Selfish* (see Sect. 5.1.1) and lottery *Considerate* typically yields the equal payoff allocation, the number of observations for the combination of kind intentions and unfair outcomes (scenario CU) is low. Accordingly, we demonstrate the causal effect of unkind intentions (unequal outcomes) by focusing on punishment differences in Scenarios SE and CE (SE and SU). As a consequence, we cannot test whether the general and potentially gender-specific effects of unkind intentions depend on the exact outcome level or the general and potentially gender-specific effect of unfair outcomes depends on the degree of kindness.

The aggregate effect of *unkind intentions* and *unequal outcomes* on punishment is illustrated in the left panel of Fig. 1. Player B's punishment was significantly higher when Player A chose *Selfish* instead of *Considerate* (78.86 points SE vs. 10.67 points CE; $p < 0.01$, WRT). This provides evidence of a significant and economically relevant effect of unkind intentions on punishment. In contrast, subjects did not choose higher punishment levels when the *unequal* instead of the *equal* outcome applies (75.73 points SU vs. 78.86 points SE; $p = 0.610$, WRT). Accordingly, in the aggregate, we do not find evidence of a significant impact of outcomes on punishment. Our first main result reads:

Result 1 *Unkind instead of kind intentions cause an increase in punishment. There is no significant impact of unequal instead of equal outcomes on punishment.*

[5] The Belief Incremental Punishment variable was constructed as follows: using the elicited punishment belief for every combination of lottery choice and payoff allocation, we first calculated the expected punishment for both lottery options and then created the expected incremental punishment variable as the difference between the expected punishment given *Selfish* and the expected punishment given *Considerate*.

[6] The insignificance results when combining all scenarios (56.8 in IMMEDIATE vs. 62.5 in DELAY, $p = 0.769$, Wilcoxon Rank-Sum Test (WRT)), when we consider single scenarios ($p > 0.34$, WRT), and when we consider either only women ($p = 0.587$, WRT) or only men ($p = 0.387$, WRT).

[7] One might argue that completing the questionnaire has, in addition to a cooling-off effect, also reduced subjects' mental energy and thereby their ability or willingness to exert cognitive effort. In contrast to this conjecture, however, subjects' earnings in the subsequent norm task, where the success probability should be positively affected by cognitive effort, are not treatment-specific ($p = 0.123$, WRT).

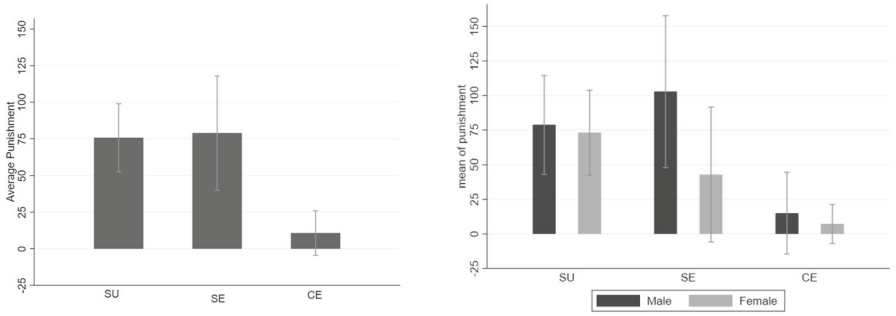

Fig. 1 Left panel: Average punishment choices by scenario. Right panel: Average punishment choices by gender and scenario

Our findings provide strong evidence for a significant intention effect on punishment, thus supporting Hypothesis 1 (b). This confirms that intentions-based reciprocity plays a crucial role in driving punishment decisions. Regarding the lack of a significant outcome effect, our results suggest that the income and inequity effects offset each other. Compared to the draw of the equal payoff, the interim payoff for Player B is substantially reduced (by 80%) following an unequal outcome. Even though the absolute level of punishment remains unchanged, Players B allocate a significantly higher proportion of their interim payoff towards punishment in response to increased inequity. This finding aligns with previous research demonstrating the significant role of inequity aversion in driving second-party punishment (e.g., Leibbrandt and Lopez-Perez 2012).

To speak about gender differences, we separate the average punishment levels by gender in the right panel of Fig. 1. Player A's choice of *Selfish* instead of *Considerate*, given the equal payoff allocation, increases males' punishment by 87.86 points ($p < 0.01$, WRT). In contrast, women's punishment increases only by 35.66 points ($p = 0.090$, WRT). Apparently, men respond more strongly to Player A's selfish lottery choice than women. Next, we turn to potential gender differences in Player B's response to the unequal payoff allocation. Men's punishment conditional on unkind intentions is 24.11 points lower in SU than in SE, a difference that is not statistically significant ($p = 0.198$, WRT). In contrast, women's punishment in SU is 30.23 points higher than in SE, which is again not statistically significant ($p = 0.343$). Although each gender's reaction is insignificant, the hypothesized gender effect might still exist as women and men change their behavior in opposite directions.[8]

Our results from non-parametric tests are confirmed in ordinary least squares regressions (Table 1). In Columns (1) and (2), the dependent variable is the punishment level, whereas it is a dummy variable equal to one when positive punishment was selected in Columns (3) and (4). The interaction of the dummy variables for the selfish lottery

[8] Given the opposing responses of women and men, a significant reaction from either gender would suffice to demonstrate differential responses to unequal payoffs. A power analysis indicates that a sample size of roughly 800 subjects in the role of Player B is necessary to detect a significant outcome effect for women. However, Table 1 leverages the contrasting responses of women and men by employing regression analyses with interaction terms. Despite our small sample size, this approach provides evidence of a gender-specific outcome effect.

Table 1 Determinants of punishment levels and incidence

	(1) Punishment Level	(2) Punishment Level	(3) Punishment Dummy	(4) Punishment Dummy
Unkind Intention	85.44***	87.69***	0.431***	0.426***
	(23.55)	(24.67)	(0.0923)	(0.0986)
Female	−3.630	−7.866	−0.0179	−0.0366
	(17.09)	(19.12)	(0.0654)	(0.0728)
Unkind Intention x Female	−64.85*	−64.82*	−0.296**	−0.288*
	(37.14)	(37.54)	(0.147)	(0.152)
Unequal Payoffs	−22.42	−28.54	−0.181*	−0.193*
	(27.04)	(28.09)	(0.107)	(0.111)
Unequal Payoffs x Female	64.76	74.95*	0.344**	0.368**
	(39.52)	(39.91)	(0.159)	(0.163)
Constant	16.24	−144.2**	0.0719	−0.301
	(13.61)	(56.41)	(0.0471)	(0.239)
Controls	No	Yes	No	Yes
N	190	190	190	190
R^2	0.084	0.120	0.114	0.129

Notes: Analysis of punishment levels (Columns (1) & (2)) and punishment dummy (= 1 if positive punishment was chosen; Columns (3) & (4)). We report results from ordinary least squares regressions. *Unkind Intention* is a dummy variable equal to one when Player A chose lottery *Selfish*. *Unequal Payoffs* is a dummy variable equal to one when the payoff allocation (1350, 150) was drawn. Controls include age, number of siblings, social value orientation, justice sensitivity, and rule-following propensity. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

choice and female is negative and at least weakly significant in every specification. The magnitude of the coefficients indicates the economic significance of the gender effect. The interaction of the dummy variables for unequal payoffs and female is positive and significant in Columns (2)-(4). Accordingly, as a reaction to a draw of unequal payoffs, women raise their punishment and are more likely than men to choose a positive punishment level. Comparing the results in Columns (1) and (3), we find that both interaction terms' significance levels are higher for the binary punishment decision than for punishment levels. This indicates that the gender effects emerge mainly due to a change at the extensive margin (the decision whether or not to punish): a draw of the unequal payoff motivates females more strongly than males to punish A, whereas the choice of *Selfish* more strongly prompts males to punish A than females. We summarize our results regarding a gender-specific relative importance of intentions and outcomes for punishment as follows:

Result 2 *(a) After a draw of the equal payoff allocation, men assign greater incremental punishment than women for Player A's choice of the selfish lottery instead of the considerate one. (b) After a selfish lottery choice of Player A, women assign greater incremental punishment than men when the unequal outcome resulted instead of the equal one.*

Table 2 Punishment levels: latent class analysis for men and women

	Punishment Level			
	Men		Women	
	Class 1 (1)	Class 2 (2)	Class 1 (3)	Class 2 (4)
Unkind Intention	6.53	286.11***	3.01	−8.42
	(13.16)	(20.82)	(10.19)	(23.90)
Unequal Payoffs	−22.50**	−10.84	10.18	75.58***
	(11.23)	(18.07)	(9.67)	(18.68)
Constant	21.20**	2.28	−0.783	202.21***
	(10.65)	(16.50)	(6.08)	(23.67)
Latent Class Marginal Probabilities	72%	28%	81%	19%

Notes: Analysis of punishment levels using GSEM regression. *Unkind Intention* is a dummy variable equal to one when Player A chose lottery *Selfish*. *Unequal Payoffs* is a dummy variable equal to one when the payoff allocation (1350, 150) was drawn. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Our findings indicate that, on average, women exhibit greater concern for equitable payoffs than men, who prioritize kind intentions. To investigate the prevalence and characteristics of distinct types within each gender, we conducted a latent class analysis. This analysis aimed to determine (i) the existence and proportion of different player types within each gender and (ii) how these types differentially respond to unequal payoffs and unkind intentions in their punishment decisions.

Table 2 reveals distinct player types for each gender. Both men and women exhibit a prevalent Class 1 type that is largely indifferent to unkind intentions and unequal payoffs, exhibiting no increase in punishment in response to either. Notably, within this class, men tend to decrease punishment following unequal payoffs, potentially driven by an income effect. This type, characterized by low punishment levels, aligns with a narrowly self-interested decision-making style. Crucially, gender-specific types emerge. Men display a second type that strongly punishes unkind intentions while demonstrating indifference to unequal payoffs. Conversely, the second type among women prioritizes punishing unequal payoffs while exhibiting little concern for intentions. Our structural estimation results thus highlight a striking gender disparity. Within our sample and experimental design, only women demonstrate a propensity to punish unequal payoffs, while only men tend to punish unkind intentions.

5.2 Social norms (part 2)

When eliciting social appropriateness ratings, we informed each subject about the gender of the subject whose norm rating they must match to obtain additional payment. However, none of the 24 punishment ratings depends on the announced gender of the paired subject ($p > 0.150$ for women and $p > 0.237$ for men). Regarding the appropriateness of lottery choices, ratings do not depend on the matched subject's gender except that women rate the *Selfish* choice as more appropriate when matched

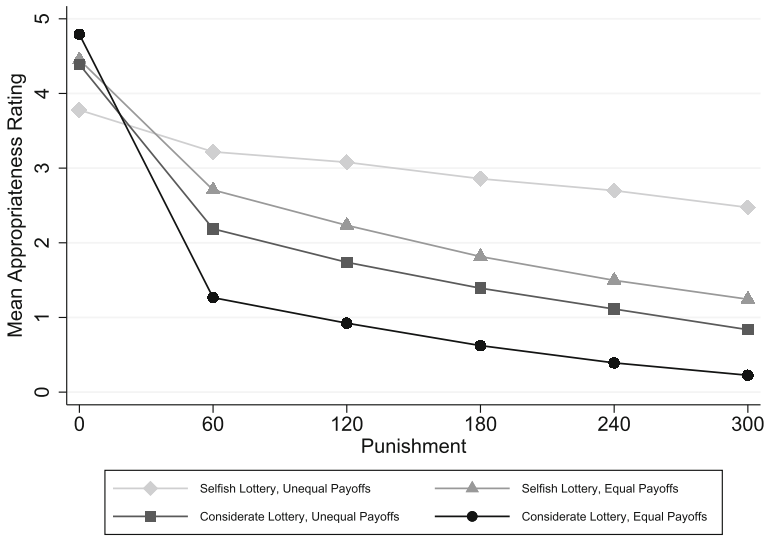

Fig. 2 Mean norm ratings of punishment in scenarios SU, SE, CU, and CE

with a man ($p = 0.064$). This suggests that any gender differences in social norms are not commonly known. In our analysis, we pool the data of same-sex and mixed-sex pairs.

5.2.1 Punishment

We identify how unkind intentions and the unequal outcome influenced the normative evaluation of punishment before we explore potential gender differences.

Zero punishment receives the highest average appropriateness rating in all scenarios (Fig. 2).[9] This speaks to the question recently raised by Fehr et al. (2018) about whether a social norm of punishment exists. The social appropriateness of punishment strongly depends on Player A's lottery choice and the drawn payoff allocation (Fig. 2). Below, we state that punishment is more appropriate in Scenario X than Y if positive punishment levels are more and zero punishment is less appropriate in Scenario X than in Y.

Independent of the outcome, punishment is more socially appropriate when Player A's intentions were unkind instead of kind ($p < 0.0001$ for every comparison, WSR). Hence, Player A's choice of *Selfish* promotes punishment. Conditional on the (un)kind intention of Player A, punishment is more socially appropriate when the unequal instead of the equal outcome was drawn ($p < 0.0001$ for every comparison, WSR). Thus, inequity legitimizes punishment. These results imply that punishment is least (most) appropriate in Scenario CE (SU). Comparing Scenarios SE and CU (i.e., scenarios with intentions and outcomes of opposite valence), we find that zero punishment

[9] There is some heterogeneity in this regard at the subject level. For some subjects, the maximal appropriateness rating applies to a positive punishment level (at least in some scenarios).

Table 3 Determinants of punishment levels and incidence conditional on norm information

	Punishment Level			
	(1)	(2)	(3)	(4)
Unequal Payoffs	3.520	−8.941	4.793	−7.145
	(20.23)	(19.57)	(20.73)	(19.94)
Unkind Intention	59.37***	37.11*	62.64***	39.66**
	(19.12)	(19.78)	(19.20)	(19.91)
Inappropriateness of Punishment		−10.05***		−9.718***
		(3.008)		(3.017)
Constant	14.53*	73.27***	−112.8**	−37.42
	(8.568)	(20.20)	(53.09)	(60.21)
Controls	No	No	Yes	Yes
N	190	190	190	190
R^2	0.063	0.147	0.089	0.166

Notes: Analysis of punishment levels. We report results from ordinary least squares regressions. *Unkind Intention* is a dummy variable equal to one when Player A chose the lottery *Selfish*. *Unequal Payoffs* is a dummy variable equal to one when the payoff allocation (1350, 150) was drawn. The variable *Inappropriateness of Punishment* is the difference in appropriateness rating between zero and maximum punishment for the relevant scenario. Controls include age, number of siblings, social value orientation, justice sensitivity, and rule-following propensity. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

is considered equally appropriate in both cases ($p = 0.834$, WSR). In contrast, all positive punishment levels are significantly more appropriate in SE ($p < 0.0001$, WSR). This demonstrates that unkind intentions increase the appropriateness of punishment by more than a draw of unequal payoffs and resonates well with observed punishment choices.

We summarize our results regarding the *averaged* social norm of punishment as follows:

Result 3 *(a) Punishment is more socially appropriate if intentions are unkind instead of kind, and if outcomes are unequal instead of equal. (b) Unkind intentions increase the social appropriateness of punishment by more than unequal payoffs, relative to a scenario with kind intentions and equal payoffs.*

To assess the explanatory power of social norms regarding punishment, we exploit individual heterogeneity regarding the inappropriateness of punishment. Controlling for Player A's intentions and the drawn outcome, Players B who rated punishment as more inappropriate should punish Player A less. The regression results displayed in Columns (2) and (4) of Table 3 confirm this prediction: the coefficient *Inappropriateness of Punishment*, which is calculated as the difference in appropriateness ratings between zero and maximum punishment, is negative and highly significant. Strikingly, the *Unkind Intention* coefficient, which captures the impact of unkind intentions, becomes smaller and less significant. A two-tailed t-test on the equality of the *Unkind Intention* coefficients in Columns (1) and (2) (and in Columns (3) and (4)) reveals that the difference of coefficients is highly significant ($p < 0.01$). Hence,

the effect of unkind intentions on punishment can (at least partly) be explained by a preference for norm compliance.

Next, we analyze gender differences in norm ratings. First, we aim to understand whether the impact of unkind intentions on the perceived social appropriateness of punishment is different for women and men. We run regressions with the appropriateness rating of punishment as the dependent variable. As independent variables, we consider the gender of the rater, Player A's intention, and whether a zero or a positive punishment level was rated. Fixing the outcome (to E in Column (1) and to U in Column (2)), the coefficient of the triple interaction shows that, for women, unkind intentions are associated with a smaller increase in the appropriateness ratings of positive punishment levels relative to zero punishment than for men. In other words, men's perceived social norm of punishment is more strongly affected by a change in intentions than the corresponding perception of women. Likewise, we explore the implications of the outcome, fixing Player A's lottery choice (to Considerate in Column (3) and Selfish in Column (4)). The coefficient of the triple interaction is insignificant in (3) and only weakly significant in (4). Thus, the impact of the outcome draw on the perceived social norm of punishment seems not to be gender-specific.

Second, we evaluate whether gender differences in perceived social norms can help to explain gender differences in punishment. Table 5 presents results from augmenting the empirical model from Table 1 by incorporating individual appropriateness ratings. The interaction of *Selfish* and *Female* becomes insignificant in our analyses of punishment levels (Columns (1) and (2)). In contrast, the gender-specific punishment response to unequal payoffs cannot be similarly explained by heterogeneity in norm ratings. The interaction of *Unequal Payoffs* and *Female* remains significant in our analysis of punishment levels. In sum, gender-specific punishment norms (i) can help to explain the gender effect in terms of punishing unkind intentions, but (ii) have little explanatory power regarding the gender-specific outcome effect.[10]

Result 4 *Incorporating social-norm ratings at the subject level helps to explain the (general as well as the gender-specific) impact of intentions on punishment choices.*

5.2.2 Lottery choice

Selfish is perceived as significantly less socially appropriate than *Considerate* (Fig. 3). Accordingly, A's choice of *Selfish* is a clear norm violation. This holds for both genders ($p < 0.01$, WRT). It is well established that norm violations are frequently punished (Fehr and Fischbacher 2004) and that such punishments help to sustain cooperation. Accordingly, Players B might have punished A for violating the lottery choice norm. This motive could be reflected in Player B's punishment norm: this is true if A's norm violation renders B's punishment more appropriate. It could, however, also operate independently from punishment norms: one possibility would be that A's norm violation triggers a socially inappropriate retaliation motive in B.

[10] Accordingly, we have tested two potential explanations for the gender-specific punishment response to unequal payoffs (emotions and social norms) and have to reject both. Hence, further research is needed to explain this result.

Table 4 Determinants of the appropriateness ratings for punishment levels

	Punishment Appropriateness Rating			
	(1) SE vs. CE	(2) SU vs. CU	(3) CE vs. CU	(4) SE vs. SU
Female	0.130	0.510***	0.130	0.296*
	(0.114)	(0.169)	(0.114)	(0.151)
Punishment Dummy	−4.864***	−3.365***	−4.864***	−2.707***
	(0.139)	(0.221)	(0.139)	(0.193)
Female x Punishment Dummy	−0.126	−0.318	−0.126	−0.717***
	(0.191)	(0.281)	(0.191)	(0.257)
Unkind Intention	−0.490***	−0.898***		
	(0.126)	(0.164)		
Female x Unkind Intention	0.166	0.333		
	(0.158)	(0.209)		
Unkind Intention x Punishment Dummy	2.157***	2.856***		
	(0.207)	(0.235)		
Female x Unkind Intention x Punishment Dummy	−0.591**	−0.862***		
	(0.272)	(0.310)		
Unequal Payoffs			−0.672***	−1.081***
			(0.148)	(0.139)
Female x Unequal Payoffs			0.380**	0.547***
			(0.175)	(0.185)
Unequal Payoffs x Punishment Dummy			1.499***	2.198***
			(0.216)	(0.203)
Female x Unequal Payoffs x Punishment Dummy			−0.191	−0.463*
			(0.265)	(0.270)
Constant	5.686***	5.014***	5.686***	5.196***
	(0.0919)	(0.137)	(0.0919)	(0.122)
N	4560	4560	4560	4560
R^2	0.489	0.274	0.488	0.205

Notes: Analysis of appropriateness ratings for different punishment levels. We report results from ordinary least squares regressions. *Unkind Intention* (*Unequal Payoffs*) is a dummy variable equal to one when Player A chose the lottery *Selfish* (when the payoff allocation (1350, 150) was drawn) in the relevant scenario that the subject evaluated. *Punishment Dummy* is equal to one when the circumstance to be evaluated features a positive punishment by Player B. Standard errors (in parentheses) are clustered at the subject level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5 Determinants of punishment levels and incidence conditional on norm information

	(1) Punishment Level	(2) Punishment Level	(3) Punishment Dummy	(4) Punishment Dummy
Unkind Intention	58.88**	61.96**	0.341***	0.335***
	(24.74)	(25.46)	(0.0971)	(0.102)
Female	−0.737	−3.038	−0.00803	−0.0195
	(15.41)	(17.46)	(0.0609)	(0.0684)
Unkind Intention x Female	−54.65	−55.72	−0.261*	−0.256*
	(35.73)	(36.22)	(0.143)	(0.148)
Unequal Payoffs	−38.23	−41.96	−0.235**	−0.240**
	(27.01)	(27.81)	(0.107)	(0.111)
Unequal Payoffs x Female	68.60*	77.99**	0.357**	0.379**
	(38.48)	(39.10)	(0.156)	(0.160)
Inappropriateness	−10.07***	−9.563***	−0.0343***	−0.0339***
of Punishment	(3.081)	(3.104)	(0.0115)	(0.0116)
Constant	73.78***	−63.71	0.268***	−0.0159
	(22.61)	(62.72)	(0.0818)	(0.258)
Controls	No	Yes	No	Yes
N	190	190	190	190
R^2	0.164	0.189	0.171	0.182

Notes: Analysis of punishment levels (Columns (1) & (2)) and punishment dummy (= 1 if positive punishment was chosen; Columns (3) & (4)). We report results from ordinary least squares regressions. *Unkind Intention* is a dummy variable equal to one when Player A chose the lottery *Selfish*. *Unequal Payoffs* is a dummy variable equal to one when the payoff allocation (1350, 150) was drawn. The variable *Inappropriateness of Punishment* is the difference in appropriateness rating between zero and maximum punishment for the relevant scenario. Controls include age, number of siblings, social value orientation, justice sensitivity, and rule-following propensity. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

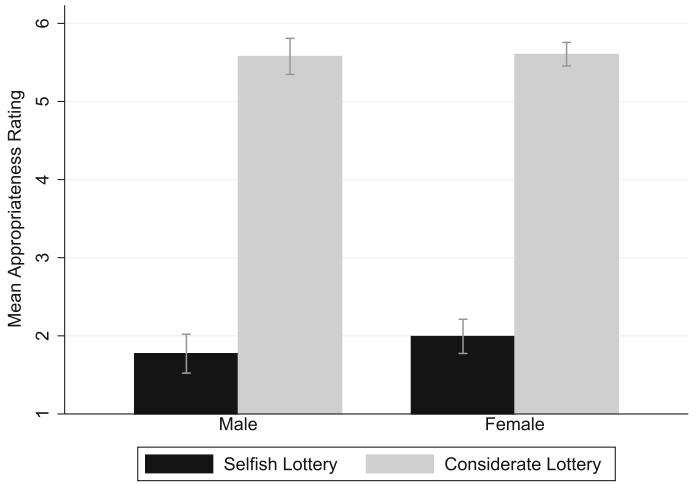

Fig. 3 Mean Norm Ratings and 95% confidence intervals of Lottery Choice by Gender

Table 6 The impact of the social appropriateness of lottery choices on punishment

	(1) SU	(2) SU	(3) SE	(4) SE	(5) CE	(6) CE
Lottery Norm	−3.161	−6.172	1.319	−4.372	−9.076**	−2.572
	(7.366)	(7.842)	(6.581)	(6.464)	(4.054)	(3.500)
Punishment Norm		−7.649		−10.51***		−35.83***
		(6.985)		(3.041)		(7.172)
Constant	90.02***	129.0**	70.46**	118.0***	48.42**	228.7***
	(32.83)	(48.34)	(28.83)	(30.66)	(18.40)	(38.98)
N	35	35	103	103	45	45
R^2	0.006	0.041	0.000	0.107	0.104	0.438

Notes: Analysis of the punishment level using ordinary least squares regressions. *Lottery Norm* is the difference in appropriateness ratings between the choices *Considerate* and *Selfish*. *Punishment Norm* is the difference in appropriateness ratings between zero and maximum punishment for the relevant scenario. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The results in Table 6 document that the punishment level is significantly correlated with the lottery norm only in CE: the more inappropriate the *Selfish* choice is, the less B punishes A when she complies with the norm by choosing *Considerate*. For all three scenarios, the coefficient for the lottery norm is insignificant when controlling for the punishment norm. In summary, our results indicate that, at least to some extent, the social appropriateness of lottery choices affects punishment norms and determines punishment choices this way. Our results do not provide evidence in favor of a separate channel, unrelated to punishment norms.

6 Conclusion

Costly punishment plays a pivotal role in fostering cooperation and promoting societal well-being (e.g., Bowles and Gintis 2004). To effectively assess and respond to behavior, individuals evaluate others' choices based on both intentions and outcomes. While previous research has explored the influence of these factors on punishment decisions, the relative importance of intentions and outcomes remains a subject of inquiry.

Our study demonstrates that unkind intentions significantly increase punishment levels at the aggregate level. In contrast, when intentions are held constant, the severity of punishment does not vary substantially with the outcome. We observe gender-specific differences in the relative importance of intentions and outcomes. Men tend to prioritize punishing unkind intentions, aligning with their preference for adhering to principles (e.g., Del Giudice et al. 2012; Eckel and Grossman 1996). In contrast, women respond more strongly to unequal outcomes, suggesting a greater concern for equal payoffs.

By examining elicited social norms, we shed light on the underlying mechanisms driving these gender differences. Our findings suggest that men and women may adhere to distinct perceptions of social norms regarding punishment.

Our results underscore the importance of considering gender-specific preferences in various domains. For example, we may consider similarities to the question about the

relative desirability of equal opportunities (i.e., a fair procedure) and similar outcomes (i.e., a fair outcome). This question is important in an organizational and a wider societal context (e.g., in terms of preferences for redistribution). Considering this reality, additional research about gender differences regarding procedural and outcome fairness using different experimental setups is warranted.

Acknowledgements We thank two anonymous reviewers for their valuable suggestions on earlier manuscript versions. In addition, we gratefully acknowledge the helpful comments received from Florian Baumann, Andreas Grunewald, Zohal Hessami, Mario Mechtel, Cat Lam Pham, Christoph Rössler, Hannes Rusch, and participants of the CESifo Area Conference on Public Economics, the seminar LIEN at Paris Nanterre, and the LawEcon Workshop at the University of Bonn.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability Data are available from the authors upon request.

Declarations

Conflict of interest None.

Appendix

A Lottery Choice of Player A

Table 7 Determinants of lottery choice

	Choice of Lottery *Selfish*		
	(1)	(2)	(3)
Female	−0.179	−0.214	−0.0214
	(0.195)	(0.201)	(0.237)
Belief Incremental Punishment		−0.00343***	−0.00371***
		(0.00111)	(0.00113)
SVO			−0.0498***
			(0.00875)
Payoff Rule-Following Task			−0.00220
			(0.00202)
Justice Sensitivity			0.0184
			(0.0853)
Age			0.0126
			(0.0123)
# Siblings			0.0851
			(0.0947)
Constant	0.691***	1.058***	2.190***
	(0.139)	(0.196)	(0.687)
N	190	190	190
Pseudo R^2	0.004	0.055	0.211

Notes: We present results from probit regressions where the dependent variable is equal to one when Player A chose lottery *Selfish*. The control variable *Belief Incremental Punishment* reflects the difference between the expected punishment contingent on *Selfish* and the expected punishment contingent on *Considerate*. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

B Treatment effect

Table 8 The role and gender-specificity of delay

	(1) Punishment Level	(2) Punishment Level	(3) Punishment Level
DELAY	13.75	12.97	14.90
	(24.24)	(23.67)	(24.97)
Female	−6.000	−6.916	−5.331
	(22.08)	(21.00)	(22.30)
DELAY X Female	−16.69	−11.62	−12.54
	(32.01)	(31.32)	(31.54)
Unkind Intention		57.77***	60.24***
		(19.39)	(19.34)
Unequal Payoffs		5.189	3.647
		(20.95)	(21.18)
Constant	60.00***	14.64	−143.5**
	(16.94)	(15.93)	(57.48)
Controls	No	No	Yes
N	190	190	190

Notes: Analysis of punishment levels. We report results from ordinary least squares regressions. *Unkind Intention* is a dummy variable equal to one when Player A chose lottery *Selfish*. *Unequal Payoffs* is a dummy variable equal to one when the payoff allocation (1350, 150) was drawn. Controls include age, number of siblings, social value orientation, justice sensitivity, and rule-following propensity. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

C Gender differences in appropriateness ratings

In the left panel of Fig. 4, we display how Player A's unkind instead of kind intentions influence the punishment levels' appropriateness ratings when the equal payoff allocation was drawn. The reported mean values stem from distributions of differences (rating of punishment level p in Scenario SE minus the rating of p in CE) at the subject level. Negative (positive) values indicate that a given punishment level is less (more) appropriate when Player A's lottery choice was *Selfish* instead of *Considerate*. For men, we find a greater positive effect on all positive punishment levels and an absolutely greater negative effect on zero punishment. In other words, circumstances with unkind intentions (de)legitimize (zero) punishment more strongly for men. The gender gap is significant for the punishment levels 180 ($p = 0.084$, WRT), 240 ($p = 0.046$), and 300 ($p = 0.013$). Differences in normative evaluations thus seem to be able to explain our result that males impose greater incremental punishment in response to Player A's selfish lottery choice. In the right panel of Fig. 4, we display the effect of

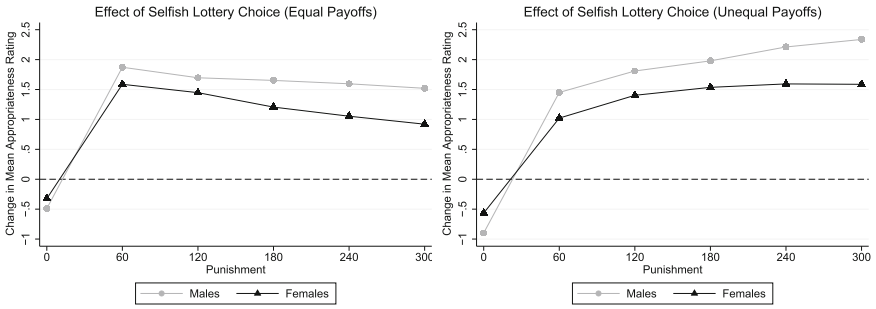

Fig. 4 Left panel: Changes in average norm ratings of punishment levels due to Player A choosing *Selfish* instead of *Considerate* given equal (unequal) payoffs in the left (right) panel

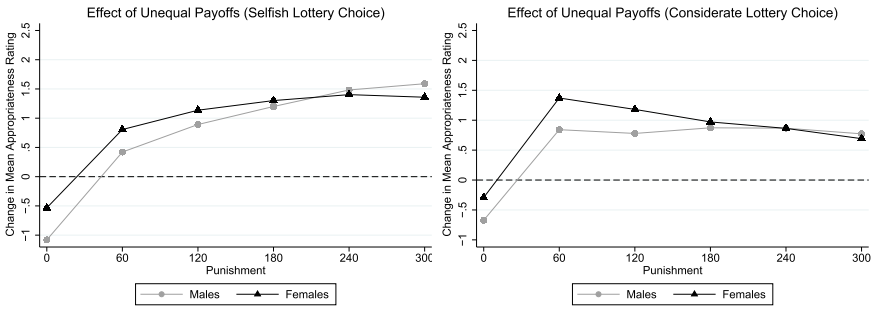

Fig. 5 Changes in average norm ratings of punishment levels due to a draw of unequal instead of equal payoffs given choice *Selfish* (*Considerate*) of Player A in the left (right) panel

unkind intentions given the equal payoff allocation. We can document a significant difference between the punishment levels of women and men for exactly the same comparisons for which we established a significant gender gap when the unequal payoff allocation applies. The positive effect of Player A's *Selfish* choice given unequal payoffs is significantly stronger for men than for women for the punishment levels 180 ($p = 0.091$, WRT), 240 ($p = 0.017$), and 300 ($p < 0.01$).

In the left panel of Fig. 5, we show how obtaining the unequal instead of the equal outcome influences the punishment levels' appropriateness ratings when the lottery choice was *Selfish*. In alignment with actual punishment choices, we find that females' appropriateness ratings increase by more than those of males for intermediate punishment levels (significant for 60, $p = 0.020$, WRT, and 120, $p = 0.069$, WRT). However, surprisingly, males' ratings of zero punishment show that men find not punishing when in scenario SU instead of SE to be much less socially appropriate ($p = 0.010$, WRT). The right panel depicts the impact of the unequal payoff draw given Player A's choice of *Considerate*. Again, the results are very similar to those depicted in the left panel. The positive effect on the social appropriateness of punishment is significantly stronger for women than for men for the punishment levels

60 ($p = 0.014$, WRT) and 120 ($p < 0.01$). The gender gap is also significant for no punishment ($p < 0.01$) with women showing a weaker negative reaction.

References

Alempaki D, Colamn AM, Kölle F, Loomes G, Pulford BD (2022) Investigating the failure to best respond in experimental games. Exp Econ 25:656–79

Andreoni J, Vesterlund L (2001) Which is the fair sex? gender differences in altruism. Q J Econ 116(1):293–312

Bagaric M (2001) Punishment and Sentencing: A Rational Approach. Cavendish, London

Balafoutas L, Nikiforakis N, Rockenbach B (2014) Direct and indirect punishment among strangers in the field. Proc Natl Acad Sci 111(45):15924–15927

Barr A, Lane T, Nosenzo D (2018) On the social inappropriateness of discrimination. J Public Econ 164:153–164

Bartling B, Engl F, Weber RA (2014) Does willful ignorance deflect punishment? an experimental study. Eur Econ Rev 70:512–24

Baumert A, Beierlein C, Schmitt M, Kemper C, Kovaleva A, Liebig S, Rammstedt B (2014) Measuring four perspectives of justice sensitivity with two items each. J Pers Assess 96(3):380–390

Bellemare C, Kröger S, Van Soest A (2008) Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. Econometrica 76(4):815–839

Blount S (1995) When social outcomes aren't fair: The effect of causal attributions on preferences. Organ Behav Hum Decis Process 63:131–144

Bock O, Baetge I, Nicklisch A (2014) hroot: Hamburg registration and organization online tool. Eur Econ Rev 71:117–120

Bolton GE, Brandts J, Ockenfels A (2005) Fair procedures: Evidence from games involving lotteries. Econ J 115(506):1054–1076

Bowles S, Gintis H (2004) The evolution of strong reciprocity: cooperation in heterogeneous populations. Theor Popul Biol 65:17–28

Brandts J, Charness G (2011) The strategy versus the direct-response method: a first survey of experimental comparisons. Exp Econ 14(3):375–398

Briton NJ, Hall JA (1995) Beliefs about female and male nonverbal communication. Sex Roles 32(1):79–90

Buchan N, Croson R, Solnick S (2008) Trust and gender: An examination of behavior and beliefs in the investment game. J Econ Behav Organ 68(3–4):466–476

Cappelen A, Drange AW, Sorensen E, Tungodden B (2007) The pluralism of fairness ideals: An experimental approach. American Economic Review 97:818–27

Cardella E, Chiu R (2012) Stackelberg in the lab: The effect of group decision making and cooling-of periods. J Econ Psychol 33(6):1070–1083

Chang D, Chen R, Krupka E (2019) Rhetoric matters: A social norms explanation for the anomaly of framing. Games Econom Behav 116:158–178

Charness G (2004) Attribution and reciprocity in a simulated labor market: An experimental investigation. Journal of Labour Economics 22:665–88

Charness G, Levine D (2007) Intention and stochastic outcomes: An experimental study. Economic Journal 117:1051–72

Charness G, Rabin M (2002) Understanding social preferences with simple tests. Quart J Econ 117:817–69

Chaudhuri A, Gangadharan L (2007) An experimental test of trust and trustworthiness. South Econ J 73:959–985

Colman AM (2006) The puzzle of cooperation. Nature 440(7085):744–745

Cox J (2004) How to identify trust and reciprocity. Games Econom Behav 46:260–281

Croson R, Buchan N (1999) Gender and culture: International experimental evidence from trust games. American Economic Review 89:386–391

Croson R, Gneezy U (2009) Gender differences in preferences. Journal of Economic literature 47(2):448–74

Cubel M, Sanchez-Pages S (2022) Gender differences in equilibrium play and strategic sophistication variability. J Econ Behav Organ 194(6):287–299

d'Adda G, Drouvelis M, Nosenzo D (2016) Norm elicitation in within-subject designs: Testing for order effects. J Behav Exp Econ 62:1–7

Dato S, Nieken P (2014) Gender differences in competition and sabotage. J Econ Behav Organ 100:64–80
Del Giudice M, Booth T, Irwing P (2012) The distance between mars and venus: Measuring global sex differences in personality. PLoS ONE 7(1):e29265
Dickinson DL, Masclet D (2015) Emotion venting and punishment in public good experiments. J Public Econ 122:55–67
Eckel C, Grossman P (1996) The relative price of fairness: gender differences in a punishment game. J Econ Behav Organ 30:143–58
Eckel CC, Grossman PJ (2001) Chivalry and solidarity in ultimatum games. Econ Inq 39(2):171–188
Eckel CC, Grossman PJ (2008) Differences in the economic decisions of men and women: experimental evidence. Handbook of Experimental Economics Results 1:509–519
Engel C (2011) Dictator games: A meta study. Exp Econ 14(4):583–610
Erkut H, Nosenzo D, Sefton M (2015) Identifying social norms using coordination games: Spectators vs. stakeholders. Econ Lett 130:28–31
Falk A, Fischbacher U (2006) A theory of reciprocity. Games Econom Behav 54:293–315
Falk A, Fehr E, Fischbacher U (2003) On the nature of fair behavior. Econ Inq 41:20–26
Falk A, Fehr E, Fischbacher U (2008) Testing theories of fairness - intentions matter. Games Econom Behav 62(1):287–303
Fehr E, Fischbacher U (2004) Third-party punishment and social norms. Evol Hum Behav 25(2):63–87
Fehr E, Gächter S (2000) Cooperation and punishment in public goods experiments. American Economic Review 90(4):980–994
Fehr E, Naef M, Schmidt KM (2006) Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment. American Economic Review 96(5):1912–1917
Fehr E, Naef M, Schurtenberger I (2018) Normative foundations of human cooperation. Nat Hum Behav 2(7):458–468
Fischbacher U (2007) z-tree: Zurich toolbox for ready-made economic experiments. Exp Econ 10(2):171–178
Friehe T, Utikal V (2018) Intentions under cover-hiding intentions is considered unfair. J Behav Exp Econ 73:11–21
Fujita F, Diener E, Sandvik E (1991) Gender differences in negative affect and well-being: the case for emotional intensity. J Pers Soc Psychol 61(3):427–434
Grimm V, Mengel F (2011) Let me sleep on it: Delay reduces rejection rates in ultimatum games. Econ Lett 111(2):113–115
Gross J, Dreu C (2021) Rule following mitigates collaborative cheating and facilitates the spreading of honesty within groups. Pers Soc Psychol Bull 47(3):395–409
Gürerk Ö, Irlenbusch B, Rockenbach B (2006) The competitive advantage of sanctioning institutions. Science 312(5770):108–111
Hopfensitz A, Reuben E (2009) The importance of emotions for the effectiveness of social punishment. Econ J 119(540):1534–1559
Joffily M, Masclet D, Noussair CN, Villeval MC (2014) Emotions, sanctions, and cooperation. South Econ J 80(4):1002–1027
Khadjavi M (2015) On the interaction of deterrence and emotions. Journal of Law, Economics, & Organization 31:287–319
Kimbrough EO, Vostroknutov A (2016) Norms make preferences social. J Eur Econ Assoc 14:608–38
Kimbrough EO, Vostroknutov A (2018) A portable method of eliciting respect for social norms. Econ Lett 168:147–150
Krupka EL, Weber RA (2013) Identifying social norms using coordination games: Why does dictator game sharing vary? J Eur Econ Assoc 11(3):495–524
Leibbrandt A, Lopez-Perez R (2012) An exploration of third and second party punishment in ten simple games. J Econ Behav Organ 84(3):753–766
Masclet D, Villeval M-C (2008) Punishment, inequality, and welfare: a public good experiment. Soc Choice Welfare 31(3):475–502
McCabe KA, Rigdon ML, Smith VL (2003) Positive reciprocity and intentions in trust games. J Econ Behav Organ 52:267–75
Murphy RO, Ackermann KA, Handgraaf M (2011) Measuring social value orientation. Judgm Decis Mak 6(8):771–781
Neo WS, Yu M, Weber RA, Gonzalez C (2013) The effects of time delay in reciprocity games. J Econ Psychol 34:20–35

Niederle M (2016) Gender. Handbook of. Exp Econ 2:481–553
Nikiforakis N, Normann H-T (2008) A comparative statics analysis of punishment in public-goods experiments. Exp Econ 11:358–69
Oechssler J, Roider A, Schmitz PW (2015) Cooling off in negotiations: Does it work? J Inst Theor Econ 171(4):565
Philippsen A, Mieth L, Buchner A, Bell R (2024) Time pressure and deliberation affect moral punishment. Sci Rep 14(1):16378
Reuben E, Winden F (2010) Fairness perceptions and prosocial emotions in the power to take. J Econ Psychol 31:908–22
Rilling JK, Sanfey AG (2011) The neuroscience of social decision-making. Annu Rev Psychol 62:23–48
Robinson MD, Johnson JT (1997) Is it emotion or is it stress? gender stereotypes and the perception of subjective experience. Sex Roles 36(3):235–258
Schlag KH, Tremewan J, Weele JJ (2015) A penny for your thoughts: A survey of methods for eliciting beliefs. Exp Econ 18:457–90
Schmitt M, Baumert A, Gollwitzer M, Maes J (2010) The justice sensitivity inventory: Factorial validity, location in the personality facet space, demographic pattern, and normative data. Social Justice Research 23(2–3):211–238
Simon RW, Nath LE (2004) Gender and emotion in the united states: Do men and women differ in self-reports of feelings and expressive behavior? Am J Sociol 109(5):1137–1176
Timmers M, Fischer A, Manstead A (2003) Ability versus vulnerability: Beliefs about men's and women's emotional behaviour. Cogn Emot 17(1):41–63
Trautmann ST (2009) A tractable model of process fairness under risk. J Econ Psychol 30(5):803–813
van't Wout M, Kahn R, Sanfey AG, Aleman A (2006) Affective state and decision-making in the ultimatum game. Exp Brain Res 169:564–568
Xiao E, Houser D (2005) Emotion expression in human punishment behavior. Proc Natl Acad Sci 102(20):7398–7401