Department of Economics

University of Reading

The Dark Side of Human Decision-Making: A Review of Behavior in the Joy-of-Destruction Experiments

by Joo Young Jeon and Shakun Mago

Discussion Paper No. 2026-05

Department of Economics
University of Reading
Whiteknights
Reading
RG6 6AA
United Kingdom

www.reading.ac.uk

The Dark Side of Human Decision-Making: A Review of Behavior in the Joy-of-Destruction Experiments

Joo Young Jeon
Department of Economics, University of Reading, Reading RG6 6AA, UK. Email: j.jeon@reading.ac.uk

Shakun Mago
Robins School of Business, University of Richmond, Richmond, VA 23173, USA. Email: sdatta@richmond.edu

Date: 07 May 2026

Abstract

Despite extensive evidence on cooperation and prosocial behavior, individuals frequently engage in costly destructive actions that yield no material benefit. This paper surveys experimental literature on the Joy-of-Destruction game, a canonical framework designed to isolate such antisocial behavior. We review evidence across a range of experimental designs, treatment conditions, and subject populations, and show that destruction persists even when standard strategic, distributive, and reciprocity-based motives are eliminated. The findings suggest that such behavior is driven by a combination of intrinsic utility from harming others, belief-driven preemptive responses, and contextual pressures. Psychological traits, group identity, and environmental factors further shape destructive choices. We also examine institutional interventions that have been shown to mitigate such behavior. Finally, we compare the Joy-of-Destruction framework to related money-burning games to clarify how differences in experimental design shape the interpretation of destructive behavior – whether as intrinsically motivated or as a response to inequality and procedural concerns.

JEL Codes: C91; D63; D74; D91

Keywords: Joy of Destruction; Money Burning; Antisocial Behavior; Nastiness; Experiment

1. Introduction

Human behavior is not always driven by rational self-interest or pro-social motives. Oftentimes, individuals willingly harm others even when doing so yields no material gain and may impose a personal cost. Such a phenomenon is often referred to as an antisocial behavior. This disturbing tendency challenges the foundational assumptions in economics and psychology and raises the need to understand such behavior in a controlled, empirical framework. In this article, we summarize the experimental literature investigating such antisocial behavior.

Antisocial behavior assumes various forms. It can be seen as sabotage (costly action to reduce the likelihood of a rival winning in competitions, Chowdhury and Gürtler, 2015), corruption (misuse of power to gain a personal advantage, Shleifer and Vishny, 1993; Treisman, 2000), vendetta (initial act of aggression leading to socially inefficient feuds, Bolle et al., 2014), antisocial punishment and sanction in a public good game (Herrmann et al., 2008; Fatas and Mateu, 2015), or deception in a trust game (Clots Figueras et al., 2015), among others. In this article, we adopt the definition of antisocial behavior as in Karakostas et al. (2025) – the intentional destruction or reduction of others' income even in the absence of any monetary gain to the destroyer – and focus predominantly on Joy-of-Destruction (JoD) games.

JoD is an experimental game coined by Abbink and Sadrieh (2009). In this game, two participants are anonymously paired and given a binary choice: they can simultaneously decide to destroy the other participant‘s endowment or maintain the status quo. The key feature is that destruction provides no material benefit to the destroying player. The design removes all the conventional motivations for destruction: there is no material gain from destruction as in the dictator games, there is no wrongdoing that is punished as in the public goods game, and there is no reciprocity as in trust games. Since endowments are equal (in expectation), there are no fairness or inequality aversion considerations as in money-burning games (Zizzo and Fleming, 2011), whereas anonymity prevents social comparison and status seeking. Removing all these strategic aspects allows researchers to isolate and study the *intrinsic pleasure some individuals may derive from causing others harm*.

Standard economic theory predicts that a rational and strictly self-interested individual will not engage in destruction, especially if destruction is costly for the destroyer. However, a significant body of evidence shows that a nontrivial proportion of participants choose to destroy, particularly when accountability is obscured. Karakostas et al. (2025) conduct a meta-analysis

and report that individuals engage in the destruction around 30% of the time when it is available, and the amount of destruction is 20% of the maximum allowable amount.

Despite the profound societal implications of these findings and a burgeoning body of experimental evidence in JoD, the literature remains somewhat fragmented across different game variants and disciplines. To date, there is no comprehensive survey that synthesizes these disparate results to provide a 'big picture' view of the mechanisms, drivers, and inhibitors of such behavior. This paper seeks to fill that gap by offering a unified overview of the JoD framework. This synthesis aims to provide researchers and policymakers alike with a structured understanding of why destructive behavior persists, and the specific conditions under which such behavior is either amplified or suppressed.

The remainder of the paper is organized as follows. Section 2 provides a theoretical framework for the JoD game. Section 3 analyzes the structural drivers of destructive behavior, including game parameters and repetition, while Section 4 examines the role of psychological traits and group dynamics. Section 5 discusses institutional interventions and policy implications. Section 6 provides a comparative analysis of the JoD and money-burning frameworks to distinguish between intrinsic malice and responses to inequality. Section 7 concludes with directions for future research.

2. Behavioral Foundations of the Joy-of-Destruction Game

Although JoD is a simple game, the motivation for destructive behavior in this game has several behavioral components: (i) material payoff maximization, (ii) relative payoff maximization, (iii) intensive utility from destruction, (iv) extensive utility from destruction, (v) reciprocal preference based on opponent's destruction behavior. In the laboratory, researchers often use different variations of the game to focus on one or more of these components.

To incorporate a theoretical framework that can nest these behavioral components in different experimental settings, we propose a social preference utility function as follows. For player i choosing a binary destruction action $d_i \in \{0,1\}$ against player j in a JoD game, the utility U_i is:

$$U_i(d_i) = \left(E_i - c_i d_i - \delta_j d_j\right) + s\left(\pi_i, \pi_j\right) + \alpha_i d_i + \beta_i(\delta_i d_i) + r_i\left(d_j\right) d_i$$

where E_i is initial endowment, c_i is the private cost of destruction, and δ_j is the amount getting destroyed by the other player (player j). Hence, $\pi_i = \left(E_i - c_i d_i - \delta_j d_j\right)$ is the final material

payoff, and the function $s(\pi_i, \pi_j)$ is a distributive social preference function (e.g., Fehr and Schmidt, 1999) that captures the utility of the relative payoff. The parameters α_i and β_i reflect the intensive and extensive utility from destruction decision. Specifically, $\alpha_i d_i$ shows the utility gained simply by taking destructive action, whereas $\beta_i(\delta_i d_i)$ shows the utility gained by the amount of damage made to player j due to the destructive action. Finally, the function $r_i(d_j)$ reflects the reciprocal preference (e.g., Charness and Rabin, 2002) for the perceived unkindness of the other player. If player i perceives themselves as a victim of player j's prior or anticipated destruction, the marginal utility of destruction increases.

Here, we present two illustrative examples to demonstrate how this theoretical framework can be integrated into JoD experimental settings. In the sections that follow, we do not explicitly specify the corresponding parameters or functional forms for each study, though these can be readily traced back to this framework.

Abbink and Sadrieh (2009) were the first to systematically isolate the empirical evidence for the psychological components underlying destructive behavior, particularly the intrinsic joy of destruction parameter α. They introduced a multi-round JoD game where destruction was costless. Participants were randomly assigned to an Open or a Hidden treatment. In the Open treatment, the other player surely learned whether their endowment was destroyed by the decision maker, i.e., the destruction decision was ex-post perfectly observable. However, in the Hidden treatment, subjects were informed that, in addition to any destruction they chose, the other player's endowment may also be destroyed by a random draw. This 'plausible deniability' reduced the image costs associated with the act, and along with no cost for destruction it effectively allowed the α term to dominate the decision-making process. While in the Open treatment the destruction rate was only 8.5% and declined over time, they rose to nearly 40% and remained stable in the Hidden treatment. This persistent destruction in the absence of strategic or distributive motives provides strong evidence for the 'pleasure of being nasty.'

One can argue that at least part of this destruction results from retaliation or pre-emptive retaliation.[1] However, Chowdhury et al. (2026) also report an 11% destruction rate in a baseline setting without pre-emptive retaliation, where subjects face no risk of being harmed themselves ($d_j = 0$). The continued incidence of destruction in this unilateral destruction environment

[1] It bears noting that, in a multi-round interaction, the opportunity to retaliate in subsequent rounds can trigger escalation but can also serve as a deterrent. Consequently, to effectively isolate "nastiness" most studies that use the JoD game to measure antisocial behavior limit play to a single round.

suggests that at least some behavior is intrinsically motivated, consistent with individuals deriving utility from harming others.

Many subsequent studies have replicated and extended these findings. Abbink and Herrmann (2011) used a one-shot version of the game and found that 25.8% of participants engaged in destruction in the Hidden treatment, compared to just 10% in the Open treatment. The results parallel those from dictator games involving moral wiggle room (Dana et al., 2007), where participants were significantly less generous when given an excuse to act selfishly. This reinforces the notion that moral cost and social image significantly constrain destructive tendencies. As the structural parameters of the game, such as the cost of destruction (c), the possibility of the other player engaging in destruction (d_j), $or\ the$ initial distribution of endowments (E_i and E_j) change, the trade-off between these utility components shifts. The following section examines how such game variants alter the incentives for destruction and help delineate the specific motivations underlying antisocial behavior.

3. Game Variants and Motivations for Destruction

As can be observed from the theoretical framework, the weight assigned to various behavioral drivers is highly sensitive to the parameters of the game. This section reviews how variations in game architecture, such as the cost of destruction, payoff symmetry, and the frequency of interaction alter the incentives for antisocial behavior. These modifications provide a systematic basis for isolating the specific motives identified in the model, allowing for a clearer distinction between intrinsic malice, strategic preemption, and reciprocity-based responses.

3.1 Cost of Destruction and Value of Endowment

Abbink and Sadrieh (2009) assume that the decision to destroy is costless; however, one can argue that destructive behavior is price sensitive. Since destruction yields no material gain to the destroyer, rational payoff maximization would mean that the player does not pay a cost to destroy the other player's endowment. Unless, of course, the player is motivated by nastiness. Numerous studies on antisocial behavior, using different frameworks such as joy-of-destruction, first-strike, and money-burning games, have incorporated costly destruction. In almost all such scenarios, the cost of destruction to the destroyer is less than its impact on the

other person.[2] However, there is overwhelming evidence that simply introducing a cost for destruction does not eliminate destructive behavior. This suggests that intrinsic motives, such as spite or malice, drive such behavior rather than merely strategic or retaliatory incentives.

Masclet and Rebière (2023) address a recurring methodological debate in this area: whether the 'nasty' behavior observed in JoD reflects true preferences or is an artifact of the hypothetical stakes in an experimental setting. By comparing real versus hypothetical incentives, they find that while real money slightly reduces the magnitude of destruction, the qualitative responses remain unchanged. Behavior responds similarly to both the 'price' of destruction and the initial distribution of wealth across treatments. Their findings reinforce the validity the JoD framework by showing that antisocial preferences are not merely driven by hypothetical incentives but reflect preferences that are robust to real financial stakes.

3.2 Fear and Pre-Emptive Retaliation

In multi-round interactions, the threat of retaliation plays a key role. Abbink and Sadrieh (2009) and Müller et al. (2022) found that destruction can escalate across rounds. Playing the game only once eliminates the possibility of future retaliation, and most later studies that employ the JoD game to measure antisocial behavior use a one-shot framework. However, even if the interaction is not repeated, preemptive retaliation - destruction of others driven by the fear of suffering destruction from others - remains one of the primary motives for destruction. This implies that players may expect the other player to be destructive and respond by being destructive, even if they were not inclined to do so. Abbink and Herrmann (2011) observe that even in a one-round interaction, individuals who anticipated destruction were significantly more likely to engage in destruction themselves. In the Hidden treatment, 56.3% of subjects who engaged in destruction expected others to destroy their endowment; this proportion was even higher in the Open treatment (85.7%).

3.3 First-Strike Dynamics

The fear of destruction by others and pre-emptive retaliation can be significant motivators of destruction, especially when the anticipated destruction cannot be retaliated. This is evident in

[2] For instance, Abbink and Herrmann (2011) consider a one-shot JoD game with a cost-to-impact ratio of 1:5 and report that a quarter of the subjects engage in destruction in their Hidden treatment. The sole exception is Müller et al. (2022) where the cost to the destroyer and the individual impact on others in the group is identical.

the study by Abbink and de Haan (2014), which is the first-strike game, a variant of the JoD game. Players in this game are matched with the same partner for a sequence of rounds and perform individual tasks over several rounds to accumulate their earnings. Destruction is modeled as follows: in any round of the game, a player may choose to pay a cost to 'deactivate' the other player. This action not only destroys the other player's accumulated earnings and reduces their future earning potential but also renders them unable to retaliate against the destroyer. Thus, the first strike is costly (and therefore, not rational), but it is also a form of self-protection since the player who strikes first is safe from future destruction. The authors find no evidence of pure nastiness, as there are zero instances of destruction when only one player can deactivate. Most of the destruction is attributed to first-order fear (fear of a possibly nasty opponent). Specifically, when both players can deactivate the other, the destruction rate rose to 77.8% with many players engaging in preemptive destruction. Interestingly, the authors find that most of the destruction occurred in the first round itself, and that if the relationship survived the first round without destruction, the resulting trust between players significantly reduced destructive behavior in subsequent rounds.

These findings are in line with Simunovic et al. (2013), who examine a simpler version of the first-strike game. The key manipulation in their one-round game is bilateral vs. unilateral preemptive strikes. They find that in bilateral conditions, about half of the participants choose to strike preemptively, whereas only 4% engage in destruction under unilateral conditions. Furthermore, when a defense option against opponent's strike is available, most subjects who previously chose to strike opt for the defense. Both studies show that fear of destruction by others motivates pre-emptive retaliatory destruction.

3.4 Unilateral Destruction

One can argue that bilateral destruction does not capture true nastiness if destruction is motivated by fear and reciprocity. A straightforward way to alleviate the fear of destruction and the motivation for pre-emptive retaliation is to examine unilateral destruction games, in which only one player can destroy the other player's endowment (while the latter remains passive). Since the player faces no prospect of destruction, one expects destruction to cease altogether. However, the experimental results are mixed. As described above, Abbink and de Haan (2014) and Simunovic et al. (2013) find that unilateral destruction is virtually non-existent. In contrast, Kessler et al. (2012) find that 15.5% of the players exercised their unilateral right to destruction. Chowdhury et al. (2026) also find a destruction rate of 11% in their baseline treatment without

pre-emptive retaliation. This suggests that at least some destruction is ‘intrinsically pleasurable’, i.e., players can indeed destroy out of pure malice.

3.5 Group Destruction Decision

Bauer et al. (2024) provide evidence that group decision-making can amplify destructive behavior. Using incentivized JoD tasks across diverse populations – including students, adolescents, and nationally representative adult samples totaling over 10,000 participants – they document a robust ‘destructiveness shift’, whereby individuals are more likely to engage in harmful actions when decisions are made in a group context.

The main mechanism behind this result is diffused responsibility. When actions are embedded within a group and cannot be individually attributed, the psychological cost of harming others declines. Consequently, destructive behavior increases sharply (from roughly 4% to 13%), which indicates a greater willingness to violate social norms when accountability is obscured. Importantly, this increase is not driven by reciprocity, inequality aversion, or strategic motives. Instead, it reflects a weakening of internal moral constraints, including concerns for self-image. When actions become observable or require justification through group deliberation, destructive behavior declines markedly, highlighting the disciplining role of accountability.

Taken together, these findings reveal the dual role of group decision-making: anonymity and shared responsibility can amplify intrinsic destructive tendencies, while accountability-enhancing features, such as observability and deliberation, serve to restrain them.

4. Psychological and Environmental Drivers

Significant heterogeneity in destructive behavior suggests that antisocial preferences depend on individual traits and environment, in addition to the game parameters and institutional characteristics. This section examines the psychological and environmental components that affect these choices. We analyze the interplay between personality, group identity, and external stressors to identify the conditions under which destructive impulses are most likely to manifest.

4.1 Group Identity

The economic significance of social identity is formalized by Akerlof and Kranton (2010), who argue that individuals derive utility from adhering to the norms associated with their social

categories. In this framework, identity not only motivates prosociality toward 'in-group' members but also dictates behavior toward 'out-groups.' Investigating the intersection of identity and destruction is particularly compelling because of two main reasons. First, it is documented that social identity can increase negative behavior such as conflict in the laboratory as well as in the field (Chowdhury, 2021). Second, it reveals how social categorization can transform destruction from a purely spiteful act into a tool for preserving group status or punishing perceived outsiders. Understanding these dynamics is essential for identifying whether antisocial behavior is an individual psychological quirk or a systematic outcome of social fragmentation.

Chowdhury et al. (2026) investigate the influence of political identity and income inequality on destructive behavior using a modified online JoD game with over 400 US participants. In a costless, unilateral framework, participants could reduce another's income under varying conditions of information about the other participant's income and political affiliation. Results show that 11.5% engaged in pure 'nastiness' when no identifying information was available. Inequality aversion significantly increased destruction, primarily among Republicans but not among Democrats. Revealing political identity escalated destruction among both parties, but especially among Democrats, who responded strongly to counterparty identity regardless of income differences. Republicans responded to political identity only when income information was absent, suggesting that identity and inequality function as substitutes in their decision-making. Hence, the interplay between identity and inequality influences destructive behavior.

Hall and Whitt (2024) also examine political identity as a driver of destruction in the US. They test whether affective polarization leads to a willingness to reduce the earnings of political opponents even when it provides no benefit to the decision-maker. They find that 'relative status' and the desire to punish explain the destruction rates. These results are consistent with Mill and Morgan (2022), who find that American people are more likely to behave spitefully toward political opponents, especially Clinton voters toward Trump supporters. Trump voters, while generally more spiteful, did not discriminate based on political affiliation.

Bauer et al. (2018) examine how peer behavior influences adolescents' willingness to harm others, particularly when the target is an ethnic minority (Roma in Eastern Slovakia). Using incentivized JoD tasks, they find that when peers acted destructively, individuals were more than twice as likely to imitate that behavior if the target was Roma rather than a co-ethnic person. Importantly, when peers are peaceful, discrimination virtually disappears. They also

run a norms-elicitation task, and report that judgments about the appropriateness of destructive acts toward Roma are more context-sensitive than toward co-ethnics. The results suggest that ethnic hostility underlying this destructive behavior is socially contagious, especially when directed toward minorities.

Along similar lines, Chuah et al. (2019) conduct a lab-in-field experiment to examine how the long civil conflict (1976-2005) influences destructive decisions in one-shot games between Acehnese and Javanese individuals in Indonesia. They find that participants who experience conflict are more likely to engage in destructive behavior, especially when the other participant belongs to an out-group (i.e., a different ethnicity). Additionally, females with conflict exposure tend to destroy out-group participants' payoffs more than males do. Thus, past conflicts continue to erode social capital and engender destructive behavior, particularly towards out-group members.

The previous three studies feature a 'strong' group identity. In contrast, Simunovic et al. (2013) examine a 'minimal' group identity setting where participants are 'arbitrarily' assigned to in- or out-groups. They find that there is no significant difference in destruction rates, depending on whether the opponent is an in-group or out-group member. [3]

Hence, we can summarize the role of identity on JoD games as follows: A strong group identity, such as ethnicity or political affiliation, can render destructive behavior toward out-group members as less socially inappropriate, making it more pervasive, whereas a minimal group identity is insufficient to spur such destruction.

4.2 Competitive Pressure and Scarcity

In addition to group identity, Chuah et al. (2019) suggest that environmental competition can foster antisocial dispositions. Other studies find support for this hypothesis. Prediger et al. (2014) find that greater exposure to scarcity and increased competitive pressure negatively affect individuals' disposition towards others and increase their readiness to engage in harmful anti-social behavior. They employ the JoD game in a field experiment with pastoralists from southern Namibia who belong to areas that differ in resource availability but are similar to each other in terms of their political, cultural, and social backgrounds. Similar to the other studies,

[3] Note that similar non-results were obtained when minimal identity is imposed in a conflict game either in a group setting (Chowdhury et al., 2016) or in an individual setting (Chowdhury et al., 2025).

they find that 32% of the participants reduce the other player's income even when the destruction is costly to them. More importantly, they find that participants from low-yield areas were twice as likely to engage in destructive behavior (40% vs. 23.3%). Also, like Abbink and Herrmann (2011), they find evidence of pre-emptive retaliation. However, the effect is again more pronounced in resource-scarce areas: 93% of the participants in low-yield areas expecting others to destroy their endowment engaged in destruction, compared with only 69% in high-yield areas. Even among those without this expectation, destruction was higher in low-yield areas (22% vs. 7%). The authors interpret this as evidence that antisocial attitudes emerge from exposure to persistent scarcity and greater competition. These are in line with the literature from the field (Miguel et al., 2004) and the lab (Chowdhury and Moffatt, 2017; Baik et al., 2020), where resource scarcity is recorded to be a significant determinant of conflict behavior.

Similarly, Saleh (2020) also finds a positive effect of scarcity and competitive pressure on antisocial behaviors among Syrian refugees. She finds that refugees in camp districts with fewer facilities (i.e., more scarcity) are significantly more likely to destroy others' payoffs. Consistent with the results of strong group identity, there is also evidence of in-group bias: refugees are more destructive toward people from other districts in the camp than toward those from their own district.

Jauernig et al. (2016) use a two-stage game in which they run a competition in the first stage and a JoD game in the second. They also find that the competition process can induce destructive behaviors. Their results show that losers face more severe destruction by winners compared to by losers, whereas the winners are destroyed equally by the losers and the winners. This shows that it is not simply the social identity of winner/loser, but the process of competition that escalates the destructive behaviors. Jauernig and Uhl (2019) followed Jauernig et al. (2016) environment and found that antisocial behaviors under competition are driven by spite and fear of pre-emptive retaliation. Even in the absence of competition, the perception of it can trigger destructive behavior.

4.3 Desire to Influence Others

Some scholars argue that destruction stems from a broader desire to influence outcomes, rather than sheer malice. Individuals often exhibit both pro-social and antisocial tendencies, depending on context (List, 2007; Bardsley, 2008; Herrmann and Orzen, 2008; Abbink et al., 2010; Zizzo and Fleming, 2011; Sadrieh and Schröder, 2012; Zhang and Ortmann, 2016). This

seemingly contradictory preference for acts of both generosity and destruction can be resolved if we allow some individuals to gain utility from influencing others' outcomes, regardless of whether the action is positive or negative. For example, Prediger et al. (2014) find that 16% of subjects who engaged in destruction in their field experiment also made above median contributions in the public goods game. Literature on antisocial punishment also cites it as a successful individual strategy for establishing social status and receiving its benefits (Anderson and Putterman, 2006; Herrmann et al., 2008). This desire for influence may explain the coexistence of generosity and destruction within the same individuals.

Grosch and Rau (2020) demonstrate that perceived procedural unfairness can trigger destructive behavior as a signal or protest, whereby individuals employ destructive behavior to influence others' perceptions of fairness.[4] In their experiment, participants complete work tasks, and then non-earners – who are excluded by a discriminatory payment procedure that disregards their performance – are paired with earners in a JoD game. The key finding is that non-earners under the discriminatory procedure destroy significantly more than in 'fairer' non-discriminatory control treatments characterized by either performance-based or arbitrary payment assignment. Moreover, in the discriminatory regime, destruction is high for almost all non-earners regardless of their aversion to inequality or beliefs about deservedness. This suggests that procedural unfairness can trigger 'symbolic' retaliatory destructive behavior even toward uninvolved coworkers. In contrast, under non-discriminatory regimes, individual inequality aversion helps explain who engages in more destruction. Thus, unfair procedures can shift destruction from preference-driven to protest-driven behavior.

4.4 Social Norm Enforcement

Destructive behavior can also serve as a norm-enforcement mechanism that is sensitive to contextual information. Takahashi and Tanaka (2021) argue that destruction can be rationalized as a social punishment for individuals who breach societal norms and restrictions. They conducted a large-scale JoD experiment in Japan, where the government had requested temporary business closure due to COVID-19. The stores that remained open were perceived as norm-breakers, and the authors report that participants were willing to reduce donations to non-compliant businesses at a personal cost. However, when participants were provided with

[4] This is in line with the money-burning games literature: Zizzo and Oswald (2001), Zizzo (2003), Fehr (2018), Zeballos (2018), Grosch and Rau (2020), and Gangadharan et al. (2021).

mitigating context (e.g., safety compliance of the norm-breaking store or its risk of bankruptcy), the destructive behavior returned to baseline levels. Thus, destruction in this framework serves as a tool of social punishment for breaching societal norms, and its intensity depends on the specific contextual norms at play.

4.5 Personality Traits and Emotions

Personality traits play a significant role in explaining individual differences in behavior. These traits, which are stable patterns of thinking, feeling, and behaving, influence the adaptive responses to variations in circumstances or environmental conditions. A set of studies examines whether individual personality traits can explain destructive behavior in JoD games, but the results are decidedly mixed. Measuring personality traits along the Five Factor Model, Kessler et al. (2012) find that subjects who scored high on neuroticism and extraversion engaged in less destruction, whereas subjects who scored high in openness to new experiences were more destructive. Raihani and Deutchman (2017) linked Dark Triad traits to higher destruction, especially in competitive environments. They find that participants from India had higher Dark Triad scores than the U.S. participants and were more likely to engage in antisocial behavior. This finding also aligns with Prediger et al. (2014) result that antisocial behavior thrives in competitive, resource-scarce environments where harming others can elevate one‘s relative status. Abbink and de Haan (2014) report that players’ general willingness to take risks in lotteries was unrelated to their destructive behavior.

Exploring the dark side of the entrepreneurial mindset, both El Harbi et al. (2020) and Saadaoui (2024) find that individuals with stronger entrepreneurial characteristics are more likely to choose destruction. El Harbi et al. (2020) argue that an entrepreneurial mindset may come with a latent tendency to break social norms and think outside the box. Saadaoui (2024) builds on this by focusing on entrepreneurial self-efficacy and emotional states. Using an emotion induction followed by a JoD game, Saadaoui finds that positive mood can amplify destructive behavior, especially among those with high entrepreneurial self-efficacy.

However, it is important to note that emotional primes do not uniformly drive destructive behavior, and the effect depends on contextual factors and prior experiences. For instance, contrary to Saadaoui’s result of positive mood increasing destruction among entrepreneurs, Lohmann et al. (2023) report that a rise in antisociality correlates with increased depression and negative mood. They investigate how exposure to COVID-19 affects antisocial behavior

and economic preferences using panel data from Chinese university students surveyed before and after the outbreak. They find that individuals more exposed to the virus displayed significantly increased antisocial behavior, without parallel changes in risk-taking, patience, or prosociality. Along similar lines, Zeitzoff (2014) finds that the effect of anger on destruction is conditional on individuals' exposure to violence. He employs a lab-in-the-field experiment in two southern Israeli cities differing in exposure to rocket fire. After being punished by their partner, the subjects can pay to destroy part or all of that partner's income. He finds that in the case of higher exposure (Sderot) induced anger reduces punishment/destruction, while in the case of lower exposure (Ofakim), it increases punishment. Moreover, individuals with greater personal exposure to violence are more likely to engage in destruction.

4.6 Environmental Factors

A well-established body of literature in psychology and climatology suggests that environmental conditions, particularly ambient temperature, exert a significant influence on human aggression and social cohesion. Seminal research, such as the meta-analysis by Anderson (1989), shows that heat stress increases physiological arousal and irritable affect, which in turn lowers the threshold for aggressive outbursts. While traditional economic models often treat the physical environment as a neutral backdrop for decision-making, environmental volatility can directly shape antisocial preferences. A primary example of this interaction is found in the relationship between thermal stress and the impulse to destroy.

Almås et al. (2025) investigate the impact of environmental stressors on antisocial behavior by conducting lab-in-the-field experiments under varying thermal conditions. Using a JoD framework, the authors find that exposure to extreme heat – i.e., thermal stress – significantly increases the incidence of destructive choices. The study suggests that physical discomfort and environmental volatility can lower the threshold for spiteful behavior. More broadly, these findings imply that macroeconomic or climate-induced shocks may have latent behavioral costs through the amplification of antisocial preferences.

5. Mitigating Destructive Behavior

Recognizing the drivers of destructive behavior is a prerequisite for designing tools that minimize their welfare costs. This section examines the effectiveness of various institutional interventions and policy instruments in curbing destructive actions in JoD. We review the

efficacy of different governance mechanisms ranging from formal punishment and monitoring to informal social norms and transparency and identify the conditions under which antisocial impulses can be effectively mitigated.

5.1 Framing Effects

Framing the JoD game as a communal decision, rather than a private transaction, significantly reduces destruction. Most studies on JoD frame the destruction decision as a transaction where a player pays a cost to destroy a portion of the other's endowment. This framing renders the decision to destroy as a market interaction in which a person willing to pay the price can exercise their choice. In this setting, cumulative efficiency considerations, i.e., impact of the destruction to self and others, are rarely emphasized. Hence, joint framing of self and others can make these social efficiency concerns more salient. Müller et al. (2022) frame the destruction decision as incurring a cost to reduce the group's endowment. They find that emphasizing the joint ownership of common property that is subject to destruction reduces the likelihood of its destruction to almost zero. They also report that introducing a parallel activity alongside the JoD game significantly reduces destruction. This supports the idea that boredom or lack of meaningful activity may be a significant driver for destructive behavior in JoD experiments. As an external forecast, this result can be used to explain how participation in job programs reduces arrests for violent crimes (Gelber et al., 2016).

5.2 Experimenter Demand and Observation

Vorlaufer (2019) examines how different levels of anonymity affect pro-social (generosity in the Dictator Game) and anti-social (destruction in the JoD game) behavior. Conducted in rural Namibia with 480 participants, the experiment compared double-anonymous, single-anonymous, and personal-disclosure treatments. Results show that destruction rates were highest (29.8%) in single-anonymous treatment and lowest (15.8%) when participants were required to personally disclose their decisions to the experimenter. Interestingly, the ability to remain completely anonymous (where both subject-subject and subject-experimenter identities are hidden) did not significantly increase nastiness (destruction rate of 19.4%). This contradicts the notion that anonymity always reduces moral constraints. The author concludes that while experimenter observation can subtly encourage moral behavior, double anonymity is not necessary to study destructive behavior.

5.3 Social Stimuli

Blackwell and Diamond (2017) demonstrate that simple prosocial gestures, such as hugging, eliminated destructive behavior. This idea aligns with existing broader evidence that social bonding and emotional priming can suppress antisocial impulses (Baumeister and Lobbestael, 2011). Blackwell and Diamond (2017) find that 22% of participants in the No-Hug treatment engage in destructive behavior. However, hugging before the game eliminated all destructive behavior. Hugging after the first round also reduced destruction, but to a lesser extent. Destruction dropped from 23% (before hug) to 11% (after hug).

Nigus et al. (2023) use the JoD game to investigate whether market integration and exposure to competitive environments erode socially responsible behavior. They find that individuals in areas with greater market access tend to exhibit higher rates of destruction, particularly when traditional social guardrails, such as religious norms or community punishment mechanisms, are absent or weakened. These results suggest that while markets facilitate efficiency, they may also weaken the link between individual actions and social consequences, thereby potentially increasing the frequency of 'nasty' behavior in the absence of effective institutional oversight.

Zhang et al. (2020) examine the effects of religious priming on pro-social and anti-social behaviors. They find that religious primes significantly increase giving in the dictator game, especially among nonreligious individuals. However, the priming intervention does not reduce destructive behavior in the JoD game nor change expectations about whether one's partner will destroy. The lack of priming effect on destructive behavior suggests that while religious cues may be effective for promoting positive cooperation, they are less effective at suppressing negative or spiteful motives. Finally, it is interesting to note that higher self-reported religiosity is positively correlated with both the likelihood and expectations of destruction.

6. Joy-of-Destruction and Money-Burning Games

The JoD game is often considered a variant of the money-burning (MB) game. Both paradigms share a common experimental structure: subjects can incur a private cost to reduce another participant's payoff, often in anonymous, one-shot setting without strategic or reputational incentives. The key distinction between the two games lies in the structure of preferences they are designed to identify, specifically, whether destruction yields utility directly or only through its effect on payoff distributions. We elaborate on this by contrasting studies from both games.

The money-burning (MB) game, introduced by Zizzo and Oswald (2001), is a one-shot, two-stage design in which payoff destruction is embedded in an environment with deliberately induced inequality. In the first stage, participants earn their income, and some receive an exogenously assigned, undeserved gift, thereby generating a relative inequality. In the second stage, participants may choose to destroy others' earnings at a personal cost. Their results show substantial burning/destructive behavior, and they find that the decision to burn is largely insensitive to the price of destruction. Instead, destruction is strongly related to relative wealth and rank: individuals who are more undeservedly disadvantaged are more likely to engage in destruction. Accordingly, MB games interpret destruction as *instrumental* (i.e., related to distributive social preference): serving to reduce disadvantageous inequality, correct unfair outcomes, or express disapproval of unjust procedures, and the act of destruction itself is assumed to carry no intrinsic utility.

By contrast, JoD games are explicitly designed to neutralize these distributive (and often reciprocal) preference-based explanations and to allow the possibility that destruction yields direct *intrinsic utility*. For example, in the original JoD game, Abbink and Sadrieh (2009) designed participants' endowments to be equal in expectation and found that participants who received lower realizations did not exhibit greater destruction, thereby ruling out inequality-based motives. Destructive behavior in JoD games is often interpreted as reflecting intrinsically antisocial preferences, such as spite, malice, or affective satisfaction from harm.

Zizzo (2003) modified the MB game into a unilateral money-burning game, in which participants can destroy others' wealth without concern about being burned themselves. In this setting, participants tend to burn less and exhibit greater sensitivity to the cost of burning. Nonetheless, they still engage in burning that reduces income inequalities, indicating behavior aligned with rank-egalitarian preferences. This pattern closely mirrors evidence from unilateral JoD games, where eliminating the possibility of retaliation leads to a substantial reduction in destruction. However, a non-trivial minority of participants still choose to destroy even when both retaliation and inequality considerations are absent (Chowdhury et al., 2026).

Field MB studies also emphasize fairness concerns. Kebede and Zizzo (2015) conduct an MB game with Ethiopian farmers and find that destructive behavior is driven by income inequality, particularly absolute differences rather than relative differences. They also report a negative relationship between money burning and the adoption of agricultural innovations. Similarly, Zeballos (2018) reports substantial burning (55%) in a sample of Bolivian farmers, with higher-

productivity individuals more likely to be targeted, consistent with concerns about the procedural fairness underlying income inequality. Comparable JoD field studies, such as Herrmann et al. (2008) and Prediger et al. (2014), instead find that destruction is amplified by scarcity and conflict exposure even when inequality is muted, highlighting fear and hostility rather than distributive concerns. Chowdhury et al. (2026) find inequality to be a significant determinant of destruction only for Republicans, but not for Democrats.

Procedural fairness is also central in MB games that examine mobility. In a labor market setting, Dickinson et al. (2018) and Dickinson and Masclet (2019) show that individuals disadvantaged by discriminatory hiring are more likely to destroy others' earnings, particularly when productivity assignments are perceived as procedurally unfair. Gangadharan et al. (2021) examine a two-stage MB game with endogenous income generation and exogenous income types. They show that allowing upward mobility – particularly when driven by effort rather than luck – significantly reduces destruction, especially among low-income participants, highlighting the role of procedural fairness in shaping antisocial behavior. Consistent with this interpretation, Fehr (2018) finds greater destruction when income inequality is perceived as being generated unfairly in a real-effort setting.

Gangadharan et al. (2019) find that social identity attenuates inequality-driven burning. They find that while low-income participants are more likely to destroy the earnings of higher-income counterparts, making shared identity salient substantially reduces destruction towards similarly disadvantaged in-group members. High-income participants, on the other hand, do not exhibit a comparable change in behavior in response to identity cues. This underscores an asymmetric role of identity in moderating inequality-driven destruction in MB games.

Niszczota and Grützner (2025) examine antisocial behavior in the context of technology related identity, specifically targeting the use of Large Language Models (LLMs). Utilizing an MB design, they find that individuals are willing to incur personal costs to reduce the payoffs of peers who use AI to complete tasks. This destruction is not driven by traditional envy of wealth, but rather by an 'anti-shortcut identity' sentiment or procedural fairness related to perceived effort. The findings suggest that the adoption of labor-saving technologies may trigger spiteful responses from those who adhere to traditional effort norms.

In JoD games, by contrast, identity salience often increases destruction toward out-groups, even without inequality or perceived procedural fairness, as documented in Section 4.1.

Finally, Zizzo and Fleming (2011) conducted dictator and MB games consecutively and found that generosity and destruction coexist, interpreting the results as a form of expressive norm enforcement. JoD studies report a similar coexistence (e.g., Prediger et al. (2014)) but interpret it instead as a desire to influence others' outcomes or as intrinsic antisociality.

Thus, the key difference between the MB and JoD games lies not in the behavior itself but in experimental design and the interpretation of results. MB games are structured to interpret destruction as a response to inequality and unfair processes, whereas JoD games are structured to identify destruction that survives even when other explanations, including inequality, are ruled out. That said, the contrast between the two games should not be overstated. Both games rule out standard payoff maximization and instead reveal non-standard social preferences through the same observable behavior: costly destruction. Both literatures also document substantial heterogeneity: fear, expectations, and context matter in JoD games, while expressive motives and social signaling may also play a role in MB environments. Together, the two paradigms provide a unified yet layered view of destructive behavior, ranging from norm-enforcing punishment to intrinsically motivated harm. This shared structure makes MB and JoD games natural complements in the experimental study of antisocial behavior.

7. Discussion

The Joy-of-Destruction game is a two-player game in which participants simultaneously decide whether to reduce another participant's endowment at no material gain to themselves. By design, the environment removes standard motivations for harmful behavior – such as reciprocity, strategic incentives, or inequality aversion – thereby isolating destruction that cannot be explained by conventional economic preferences. This feature makes JoD a clean empirical benchmark for identifying intrinsically antisocial behavior, often described as the 'pleasure of being nasty'.

This survey reviews a growing but increasingly nuanced literature on JoD, showing that destructive behavior is both robust and highly sensitive to context. Destruction varies systematically across information structures (e.g., hidden vs. observable decisions), institutional features (e.g., unilateral vs. bilateral destruction, first-strike environments, group decision making), and subject populations. Across these settings, the evidence highlights a range of underlying drivers – including beliefs about others' behavior, fear of preemptive harm,

social identity, and environmental conditions – while also identifying institutional interventions that can mitigate destructive tendencies.

Taken together, the JoD framework reveals that destructive behavior is not an anomaly, but a persistent feature of decision-making that emerges even in the absence of material incentives. Individuals sometimes choose to harm others not out of self-interest or retaliation, but due to intrinsic motives or context-dependent triggers. While such behavior is difficult to reconcile with standard models of preferences, experimental evidence shows that it is both systematic and predictable once contextual factors are considered. Therefore, understanding the psychological and environmental roots of destructive behavior is essential for both theory and policy. Whether through institutional design, norm framing, or social engagement, reducing the conditions that enable and amplify antisocial tendencies may be critical to improving social welfare. The JoD framework offers a unique and valuable tool for continuing this inquiry.

Future research can extend the JoD framework along several important dimensions. One promising direction is to incorporate communication and cheap talk, allowing individuals to justify or signal their intentions before making destructive choices. Another extension is to introduce richer belief elicitation and learning dynamics to understand better how expectations about others' behavior shape preemptive destruction. Embedding JoD in dynamic or repeated environments with endogenous matching can further illuminate how antisocial behavior evolves over time and whether reputational forces mitigate or reinforce it. Integrating real-effort tasks would also help assess how earned versus unearned endowments interact with destructive behavior. Finally, exploring institutional interventions – such as punishment mechanisms, reward schemes, or accountability structures – can provide deeper insights into how environments can be designed to reduce antisocial tendencies.

Literature can also be advanced by investigating heterogeneity in destructive behavior across individuals. In particular, gender remains an underexplored dimension in explicitly antisocial settings. While a substantial body of work documents that women tend to display higher prosociality and greater sensitivity to experimental conditions, including pricing, framing and competition (Niederle, 2016; Chowdhury et al., 2017; Bilén et al., 2021; Mago and Razzolini, 2019), evidence from environments involving purely antisocial choices, such as the JoD game, remains limited. Consequently, the extent to which gender shapes destructive or harm-inflicting behavior is still insufficiently understood within this framework.

References

Abbink, K., & de Haan, T. (2014). Trust on the brink of Armageddon: The first-strike game. *European Economic Review, 67*, 190–196.

Abbink, K., & Herrmann, B. (2011). The moral costs of nastiness. *Economic Inquiry, 49*, 631–633.

Abbink, K., & Sadrieh, A. (2009). The pleasure of being nasty. *Economics Letters, 105*, 306–308.

Abbink, K., Brandts, J., Herrmann, B., & Orzen, H. (2010). Intergroup conflict and intra-group punishment in an experimental contest game. *American Economic Review*, 100(1), 420–447.

Akerlof, G. A., & Kranton, R. E. (2010). *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being*. Princeton University Press.

Almås, I., Auffhammer, M., Bold, T., Bolliger, I., Dembo, A., Hsiang, S. M., Kitamura, S., Miguel, E., & Pickmans, R. (2025). Destructive behaviour, judgement, and economic decision-making under thermal stress. *Economic Journal. 135*(672), 2483–2508.

Anderson, C. A. (1989). Temperature and aggression: Ubiquitous effects of heat on individual and social aggression. *Psychological Bulletin*, *106*(1), 74–96.

Anderson, C. M., & Putterman, L. (2006). Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism. *Games and Economic Behavior, 54*(1), 1–24.

Baik, K. H., Chowdhury, S. M., & Ramalingam, A. (2020). The effects of conflict budget on the intensity of conflict: An experimental investigation. *Experimental Economics, 23*(1), 240–258.

Bardsley, N. (2008). Dictator game giving: altruism or artefact? *Experimental Economics, 11*(2), 122–133.

Bauer, M., Cahlíková, J., Celik Katreniak, D., Chytilová, J., Cingl, L., & Želinský, T. (2024). Nastiness in Groups. *Journal of the European Economic Association, 22*(5), 2075–2107.

Bauer, M., Cahlikova, J., Chytilova, J., & Zelinsky, T. (2018). Social contagion of ethnic hostility. *Proceedings of the National Academy of USA, 115*, 4881–4886.

Baumeister, R. F., & Lobbestael, J. (2011). Emotions and antisocial behavior. *Journal of Forensic Psychiatry & Psychology, 22*(5), 635–649.

Bilén, D., Dreber, A., & Johannesson, M. (2021). Are women more generous than men? A meta-analysis. *Journal of the Economic Science Association, 7*(1), 1–18.

Blackwell, C., & Diamond, Z. (2017). Combating the joy of destruction with pro-social behavior. *Review of Behavioral Economics, 4*(3), 275–293.

Bolle, F., Tan, J. H., & Zizzo, D. J. (2014). Vendettas. *American Economic Journal: Microeconomics, 6*(2), 93–130.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics, 117*(3), 817-869.

Chowdhury, S. M. (2021). *The economics of identity and conflict*. In Oxford research encyclopedia of economics and finance. Oxford University Press.

Chowdhury, S. M., & Gürtler, O. (2015). Sabotage in contests: a survey. *Public Choice, 164*(1), 135–155.

Chowdhury, S. M., & Moffatt, P. G. (2017). Overbidding and heterogeneous behavior in contest experiments: A comment on the endowment effect. *Journal of Economic Surveys*, 31(2), 572-576.

Chowdhury, S. M., Jeon, J. Y., & Mago, S. (2026). Partisanship, Inequality, and Destructive Behavior: An Experiment with Republicans and Democrats. *Working paper*.

Chowdhury, S. M., Jeon, J. Y., & Saha, B. (2017). Gender differences in the giving and taking variants of the dictator game. *Southern Economic Journal, 84*(2), 474–483.

Chowdhury, S.M., Jeon, J., & Ramalingam, R. (2016). Identity and Group Conflict. *European Economic Review*, 90, 107-121.

Chowdhury, S.M., Mukherjee, A., & Sheremeta, R. (2025). In-group versus Out-group Preferences in Intergroup Conflict: An Experiment. *Journal of Public Economic Theor*y, 27(5), e70074.

Chuah, S. H., Feeny, S., Hoffman, R., & Sanjaya, M. R. (2019). Conflict, ethnicity and gender: A money-burning field experiment in Indonesia. *Economics Letters, 177*, 14–17.

Clots Figueras, I., Hernán-González, R., & Kujal, P. (2015). Information asymmetry and deception. *Frontiers in Behavioral Neuroscience, 9,* 109.

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory, 33*(1), 67–80.

Dickinson, D. L., & Masclet, D. (2019). Using ethical dilemmas to predict antisocial choices with real payoff consequences: An experimental study. *Journal of Economic Behavior and Organization, 166*, 195–215.

Dickinson, D. L., Masclet, D., & Peterle, E. (2018). Discrimination as favoritism: The private benefits and social costs of in-group favoritism in an experimental labor market. *European Economic Review, 104*, 220–236.

El Harbi, S., Grolleau, G., Sutan, A., & Ben Ticha, Z. (2020). Are individuals with entrepreneurial characteristics more likely to engage in destruction? An experimental investigation among potential 21st century entrepreneurs in Tunisia. *Bulletin of Economic Research, 72*, 33–49.

Fatas, E., & Mateu, G. (2015). Antisocial punishment in two social dilemmas. *Frontiers in Behavioral Neuroscience, 9,* 107.

Fehr, D. (2018). Is increasing inequality harmful? Experimental evidence. *Games and Economic Behavior, 107*, 123–134.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, *114*(3), 817-868.

Gangadharan, L., Grossman, P. J., & Vecci, J. (2021). Moving on up: The impact of income mobility on antisocial behaviour. *European Economic Review, 134*, 103686.

Gangadharan, L., Grossman, P. J., Molle, M. K., & Vecci, J. (2019). Impact of social identity and inequality on antisocial behaviour. *European Economic Review, 119*, 199–215.

Gelber, A., Isen, A., & Kessler, J. B. (2016). The effects of youth employment: Evidence from New York City lotteries. *Quarterly Journal of Economics, 131*(1), 423–460.

Grosch, K., & Rau, H. A. (2020). Procedural unfair wage differentials and their effects on unethical behavior. *Economic Inquiry, 58*, 1689–1706.

Hall, J., & Whitt, S. (2024). Examining affective partisan polarization through a novel behavioral experiment: The equality equivalency test in the United States (2019–2022). *Journal of Behavioral and Experimental Economics, 112*, 102245.

Herrmann, B., & Orzen, H. (2008). *The appearance of homo rivalis: Social preferences and the nature of rent seeking*. CeDEx discussion paper series No. 2008-10.

Herrmann, B., Thoni, C., & Gachter, S. (2008). Antisocial punishment across societies. *Science, 319*(5868), 1362–1367.

Jauernig, J., & Uhl, M. (2019). Spite and preemptive retaliation after tournaments. *Journal of Economic Behavior and Organization, 158*, 328–336.

Jauernig, J., Uhl, M., & Luetge, C. (2016). Competition-induced punishment of winners and losers: Who is the target? *Journal of Economic Psychology, 57*, 13–25.

Karakostas, A., Tran, E. N., & Zizzo, D. J. (2025). Experimental insights on antisocial behavior: two meta-analyses. *Experimental Economics, 28*(1), 43–74.

Kebede, B., & Zizzo, D. J. (2015). Social preferences and agricultural innovation: an experimental case study from Ethiopia. *World Development, 67*, 267–280.

Kessler, E., Ruiz-Martos, M., & Skuse, D. (2012). *Destructor game*. Department of economics Universitat Jaume I Working paper No. 2012/11.

List, J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy, 115*(3), 482–493.

Lohmann, P. M., Gsottbauer, E., You, J., & Kontoleon, A. (2023). Anti-social behaviour and economic decision-making: Panel experimental evidence in the wake of COVID-19. *Journal of Economic Behavior & Organization, 206*, 136–171.

Mago, S. D., & Razzolini, L. (2019). Best-of-five contest: An experiment on gender differences. *Journal of Economic Behavior & Organization*, *162*, 164-187.

Masclet, D., & Rebière, T. (2023). Comparing real and hypothetical incentives in giving and money burning experiments. *Annals of Economics and Statistics*, *152*, 65-102.

Miguel, E., Satyanath, S., & Sergenti, E. (2004). Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy, 112*(4), 725–753.

Mill, W., & Morgan, J. (2022). The cost of a divided America: an experimental study into destructive behavior. *Experimental Economics, 25*, 974–1001.

Müller, J., Schwieren, C., & Spitzer, F. (2022). How to prevent destruction–On the malleability of anti-social behavior. *Journal of Behavioral and Experimental Economics, 97*, 101798.

Niederle M. (2016). Gender. In: Kegel J.H., Roth A.E., eds. Handbook of Experimental Economics. 2, 481–562. Princeton University Press, Princeton.

Nigus, H. Y., Nillesen, E., Mohnen, P., & Di Falco, S. (2023). Markets and socially responsible behavior: Do punishment and religion matter? *Journal of Economic Behavior & Organization*, *209*, 572–593.

Niszczota, P., & Grützner, C. (2025). *Antisocial behavior towards large language model users: Experimental evidence*. arXiv preprint No. 2601.09772.

Prediger, S., Vollan, B., & Herrmann, B. (2014). Resource scarcity and antisocial behavior. *Journal of Public Economics, 119*, 1–9.

Raihani, N., & Deutchman, P. (2017). *Dark Triad personality traits vary across countries and predict antisocial behavior*. PsyArxiv Preprint No. 8t6k5_v1.

Saadaoui, H. (2024). Is more always better with respect to entrepreneurial self-efficacy? An experimental investigation. *Journal of Business Economics, 94*(5), 1–26.

Sadrieh, A., & Schröder, M. (2012). *The desire to influence others*. FEMM Working Paper No. 12027.

Saleh, L. (2020). *A friend or a foe in exile: An experimental analysis of punishment in the Zaatari refugee camp in Jordan*. SSRN working paper No. 3594357.

Shleifer, A., & Vishny, R. W. (1993). Corruption. *Quarterly Journal of Economics, 108*(3), 599–617.

Simunovic, D., Mifune, N., & Yamagishi, T. (2013). Preemptive strike: An experimental study of fear-based aggression. *Journal of Experimental Social Psychology, 49*(6), 1120–1123.

Takahashi, R., & Tanaka, K. (2021). Social punishment for breaching restrictions during the COVID-19 pandemic. *Economic Inquiry, 59*(4), 1467–1482.

Treisman, D. (2000). The causes of corruption: a cross-national study. *Journal of Public Economics, 76*(3), 399–457.

Vorlaufer, T. (2019). Effects of double-anonymity on pro-and anti-social behavior: Experimental evidence from a lab in the field. *Journal of Behavioral and Experimental Economics, 81*, 216–225.

Zeballos, E. (2018). Destructive actions and productivity: Experimental evidence on interpersonal comparisons among dairy farmers in Bolivia. *Journal of Behavioral and Experimental Economics, 76*, 82–94.

Zeitzoff, T. (2014). Anger, Exposure to Violence, and Intragroup Conflict: A 'Lab in the Field' Experiment in Southern Israel. *Political Psychology, 35*(3), 309–335.

Zhang, J., Brown, E., & Xie, H. (2020). Effect of religious priming in prosocial and destructive behaviour. *Pacific Economic Review, 25*(1), 47–68.

Zhang, L., & Ortmann, A. (2016). Pro-social or anti-social, or both? A within-and between-subjects study of social preferences. *Journal of Behavioral and Experimental Economics, 62*, 23–32.

Zizzo, D. J. (2003). Money burning and rank egalitarianism with random dictators. *Economics Letters, 81*, 263–266.

Zizzo, D. J., & Fleming, P. (2011). Can experimental measures of sensitivity to social pressure predict public good contribution? *Economics Letters, 111*(3), 239–242.

Zizzo, D. J., & Oswald, A. J. (2001). Are people willing to pay to reduce others' incomes? *Annales d'Economie et de Statistique, 63/64*, 39–65.