NASTINESS IN GROUPS

Michal Bauer
CERGE-EI and Charles University, Czech Republic

Jana Cahlíková
University of Bonn, Germany

Dagmara Celik Katreniak
City, University of London, UK

Julie Chytilová
Charles University and Czech Academy of Sciences, Czech Republic

Lubomír Cingl
Prague University of Economics and Business, Czech Republic

Tomáš Želinský
Technical University of Košice, Slovakia and Durham University, UK

Abstract
This paper provides evidence showing that people are more prone to engage in nasty behavior, malevolently causing financial harm to other people at own costs, when they make decisions in a group context rather than when making choices individually on their own. We establish this behavioral regularity in a series of large-scale experiments among university students, adolescents, and nationally representative samples of adults—more than ten thousand subjects in total. We test several potential mechanisms, and the results suggest that individual nasty inclinations are systematically more likely to affect behavior when decisions are made under the "cover" of a group, that is, in a group decision-context that creates a perception of diffused responsibility. (JEL: C92, C93, D01, D64, D74, D91)

The editor in charge of this paper was Romain Wacziarg.

Acknowledgments: The data collections were funded by the Max Planck Institute for Tax Law and Public Finance (Experiment 1), ERC-CZ/AV-B (ERC300851901, Experiment 2) and the Slovak Research and Development Agency (APVV-19-0329, Experiment 3). Bauer, Chytilová and Cingl acknowledge financial support from the Czech Ministry of Education, Youth, and Sports (ERC CZ program, LL2303). Cahlíková acknowledges support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC 2126/1-390838866 and through CRC TR 224 (Project A05). We thank Roland Bénabou, Gary Charness, Stefano DellaVigna, Dirk Engelmann, Armin Falk, Uri Gneezy, Shachar Kariv, Supreet Kaur, Martin Kocher, Filip Matějka, Matthias Sutter, RomainWacziarg, Jan Zápal, anonymous reviewers, and participants at various seminars and workshops for helpful comments.

E-mail: bauer@cerge-ei.cz (Bauer); jana.cahlikova@uni-bonn.de (Cahlíková); dagmara.celik@city.ac.uk, (Katreniak); chytilova@fsv.cuni.cz (Chytilová); lubomir.cingl@vse.cz (Cingl); tomas.zelinsky@tuke.sk (Želinský)

Journal of the European Economic Association 2024 0(0):1–33
https://doi.org/10.1093/jeea/jvad072

1. Introduction

Do people become nastier to other people when they act in groups than when they act individually? Since Plato's writings about the limits of democracy (trans. 1891) and later by the founding fathers of the US constitution, social scientists have been concerned about the effects on behavior when individuals act as a part of a group.[1] For example, in his seminal work on individual behavior in crowds, Le Bon (1895, p. 35) argues that individuals placed in groups lose their sense of personal responsibility and "socially deviant" inclinations may shape behavior: "*Isolated he may be a cultivated individual; in a crowd he is a barbarian—that is, a creature acting by instinct.*" In line with this concern, anecdotal evidence suggests that some people are more prone to engage in vandalism or violence against perfect strangers in collective settings such as crowds, gangs, or armed groups. Covert obstructionism and sabotage are particularly common in complex organizations and large bureaucracies (Neuman and Baron 1998; Hershcovis et al. 2007; Gangadharan, Grossman, and Vecci 2020). Some of this harmful behavior is hard to explain by pure calculated self-regard, and suggests that some people may harbor a desire to be nasty and to harm others, even at personal cost to themselves, which may get manifested in group settings.

This paper provides controlled experimental evidence testing whether individually costly tendencies to harm others are more pronounced when decisions are made in a group context than in an individual context, and explores which property of the group setting may cause such a "destructiveness shift."[2] Addressing these questions can shed light on determinants of nasty behaviors in group settings and help to identify ways to reduce the prevalence of such behaviors. In a series of five large-scale experiments (one experiment among university students, two online experiments among nationally representative samples of adults, and two pilot lab-in-field experiments among adolescents) involving more than 10,000 individuals, we elicit destructive behavior in controlled, incentivized tasks, and document that individual nasty inclinations are systematically more likely to affect behavior when decisions are made under the "cover" of a group, that is, in a group decision-context that creates a perception of greater anonymity and more diffused responsibility.

1. In Plato's opinion, democracy involves rule by irrational mobs, and for this reason he favored the rule of an enlightened individual (Allport 1968). Alexander Hamilton, James Madison, and John Jay—the first two of whom were members of the US Constitutional Convention—shared a similar concern: *"In all very numerous assemblies, of whatever character composed, passion never fails to wrest the scepter from reason. Had every Athenian citizen been a Socrates, every Athenian assembly would still have been a mob"* (Publius 1948, p. 248).

2. Earlier experiments on decision-making in group settings made progress in studying choices, in which self-interest is pitted against social welfare or moral outcomes, using a range of experimental tasks such as the Prisoners' Dilemma game, the Trust game, and the Dictator game (see excellent surveys by Charness and Sutter 2012 and Kugler, Kausel, and Kocher 2012). In this paper, we focus on the dark side of social preferences, and elicit destructive behavior that cannot be explained by self-interest, because it is privately costly, making it particularly socially dangerous.

Being nasty is a substantial step from being self-regarding. It refers to a desire to reduce another's material payoff for the sole purpose of harming that person, *without creating personal material gain and without fairness justification* (e.g. to reduce inequality or as a reciprocal response to hostility).[3] It is widely recognized that the prevalence of prosocial behaviors, at the expense of behavior guided by pure material self-interest, is important in determining a range of desirable societal outcomes—provision of public goods, participation in public life, and cooperation (e.g. Fehr and Gächter 2002; Almas, Cappelen, and Tungodden 2020). Nastiness represents another fundamental—but thus far less-studied—departure from own money-maximizing behavior (Herrmann, Thoni, and Gachter 2008; Abbink and Sadrieh 2009), because it reduces the propensity to cooperate even in situations in which mutual cooperation is an equilibrium for selfish players (Fehr, Hoff, and Kshetramade 2008). While economic agents motivated purely by self-interest are predicted to reduce social welfare only when they can personally gain from doing so, the scope for welfare losses and mutually destructive conflicts is magnified when decision-makers derive utility from harming others.

To elicit nasty behavior, we employ a controlled money-burning task, often referred to as the Joy of Destruction mini-game (Abbink and Sadrieh 2009; Abbink and Herrmann 2011). Subjects choose whether to pay a small amount of money to lessen the reward of an anonymous counterpart. The dominant strategy for purely self-regarding individuals is to not engage in destructive behavior. Further, the destructive action cannot be explained by aversion to inequality or negative reciprocity. Therefore, the task allows us to cleanly identify the dark side of human social preferences, and to distinguish whether a group context fosters behavior motivated by a desire to be nasty that goes beyond standard own-money-maximizing selfishness.

In Experiment 1 (E.1), we study the behavior of 795 university students from Slovakia and aim to separate the roles of different mechanisms that can affect the prevalence of nasty behavior in groups. Our main focus is on the contextual features that can create a perception of group "cover." This is motivated by models that highlight the importance of self-image considerations (Benabou and Tirole 2011; Benabou, Falk, and Tirole 2018) and predict that, in decision contexts that reduce the salience of self, increase the perception of diffused responsibility, and enhance the scope for "plausible deniability" (Darley and Latane 1968; Dana, Weber, and Kuang 2006), self-image considerations become a less powerful regulator of individual social behavior. Consequently, people with a latent intrinsic preference to be nasty feel fewer scruples against violating social norms and act on such desires.

To identify the effects of group "cover," we compare choices made in the *Individual* condition, when subjects make choices on their own behalf, with choices in the *GroupContext_Hidden* (*GC_Hidden*) condition, in which subjects are matched with two other individuals, each group member makes a decision knowing that no-one will

3. The literature uses various terms when describing the dark side of social preferences, including nastiness, spitefulness, anti-social behavior, aggressive competitiveness, and destructiveness. In this paper, we refer to nastiness or destructiveness and use the two terms interchangeably.

learn which of the three individual decisions was randomly selected (with a chance of one third) and treated as the decision of the whole group. This decision-making context is designed to create a perception of diffused responsibility, because multiple individuals are involved in the decision-making and can be attributed the responsibility for the group decision. To eliminate the role of other mechanisms, such as social image considerations, identification with the group, or an expectation of future group deliberation, the group members were completely anonymous to each other, did not observe each other's decisions, and did not communicate or otherwise interact in any way.

We implement two further conditions to study two additional common features of decision-making in groups, to test whether they represent an opposing mechanism to the group cover effect, or whether they may reinforce it. First, people may care about their social image and what others think of them. This may reduce nastiness, if group members prefer to present themselves as being kind.[4] To shed light on this effect, we compare choices in *GC_Hidden* with choices in *GroupContext_Observed* (*GC_Observed)*, which differs from *GC_Hidden*, in that the subjects sat next to each other and knew that their choices would be observed by the other group members. Second, existing experiments have shown that groups often behave more in line with narrow self-interested predictions than individuals, when group members can communicate and must reach a joint decision (Charness and Sutter 2012; Kugler, Kausel, and Kocher 2012; Kocher, Schudy, and Spantig 2018). Group decisions have also been shown to lead to fewer mistakes and erroneous decisions (Charness, Karni, and Levin 2010). This evidence suggests that group communication could reduce nastiness in the Joy of Destruction game (JDG), because destruction reduces both own payoff and total welfare. To study this, we implemented a *Group* condition, in which unitary groups of three members were asked to deliberate and to make a joint decision. Comparing choices made in *Group* with choices made in *GC_Hidden and GC_Observed* allows us to estimate the effect of group communication and preference-aggregation.

We report the following main observations. First, making decisions in *GC_Hidden* substantially increases the prevalence of destructive behavior, compared to *Individual*, from 4% to 13%. Second, the prevalence of destructive behavior in *GC_Observed* (7%) is lower than in *GC_Hidden*, suggesting that observability among group members might attenuate the "destructiveness shift." In addition, the prevalence of destructiveness drops further to 2%, when group members communicate and need to reach a joint decision.

The first finding suggests that increased perceptions of anonymity and diffused responsibility in a group setting makes people more prone to violating social norms and engaging in nasty behavior, compared to when they act individually on their own behalf. Furthermore, properties of decision-making in groups that arguably reduce perceptions of group "cover," such as observability of choices by group members or the need to justify one's own views during group communication, reduce the

4. In principle, the effect could be opposite if subjects aimed to signal toughness by being nasty, however, the existing evidence suggests that most people care about having a positive social image (Andreoni and Bernheim 2009; DellaVigna, List, and Malmendier 2012).

prevalence of destructive behavior. Several alternative mechanisms that could explain the "destructiveness shift" in *GC_Hidden*, such as motivation to please nasty group members, the influence of in-group versus out-group biases, and the action bias, are not supported by our experiments.

As a next step, in Experiment 2 (E.2), we replicate the main pattern (greater prevalence of destructive behavior in *GC_Hidden* than in *Individual*) among a nationally representative online sample of adults in Slovakia ($N = 4{,}243$). Further, the large sample size allows us to document that the effect holds for various sub-groups, considering gender, age, education, income level, size of municipality of residence, and political preferences of the respondents.

In Experiment 3 (E.3), implemented among a different sample of adults ($N = 3{,}349$), we address the concern that the decision situation in JDG may prime conflict and potentially lead to experimenter demand effects. We adapt the task so that it includes an option to act kindly in addition to the option to act destructively. Thus, this modified JDG more closely reflects many real-life situations in which people have an option to be helpful or harmful. Further, the counterpart is passive and thus cannot be seen as an adversary. Importantly, we observe the same pattern as in E.1 and E.2. Subjects are more likely to choose the destructive option in *GC_Hidden* than in *Individual*, suggesting that the main effect is not driven by greater sensitivity to cues of conflict in the group setting. Interestingly, we also find that the overall level of destructiveness is generally lower in E.3 (modified JDG) than in E.2 (standard JDG), despite the fact that the sampling was comparable. Although alternative explanations are possible, this pattern suggests that the (often surprisingly high) levels of destructiveness observed in the standard JDG, in which subjects can only harm, should be interpreted with caution, and that using the version with expanded choice space might provide a more accurate picture about the prevalence of destructive preferences.

Before Experiments 1–3, we conducted two pilot experiments with adolescents in Slovakia (PE.1, $N = 630$) and Uganda (PE.2, $N = 1{,}679$). We used a somewhat different approach to elicit preferences made under group cover, asking subjects to privately state their individual preference about how they wanted their group to decide when making a joint decision in JDG, prior to the group discussion. In both countries, we find significantly higher prevalence of destructive behavior in these choices than in the *Individual*. These findings are in line with the results from the three main experiments, indicating that the observed behavioral pattern also holds among adolescents and is robust across two very different cultures.

This paper is related to several streams of literature. First, it adds to a growing literature that studies determinants of anti-social behavior in controlled experiments (Herrmann, Thoni, and Gächter 2008; Abbink and Sadrieh 2009; Abbink and Herrmann 2011; Prediger, Vollan, and Herrman 2014).[5] It contributes by providing the

5. For recent reviews of the experimental work on anti-social behavior, see Karakostas, Tran, and Zizzo (2022); Sanjaya (2023). In addition to our focus on the role of group settings, another unusual feature of our

first experimental evidence testing the role of different properties of a group setting. Our main finding relates to the literature on "moral wiggle room" and perceptions of individual responsibility. Laboratory studies have documented that contextual factors that provide more scope for excuses by obscuring the role of the decision-maker in determining an outcome (Dana, Weber, and Kuang 2006, and Weber 2011), or that reduce the salience of self (Falk 2021) increase the prevalence of unfair or immoral behavior for one's own financial benefit. Since decisions in these experiments are anonymous, they can be interpreted through the lens of a behavioral model with self-image concerns (Benabou and Tirole 2011; Benabou, Falk, and Tirole 2018), in which reduced perception of individual responsibility allows individuals to maintain a positive self-image even when they violate social norms (Krupka and Weber 2013). While much of the literature focuses on the role of self-image concerns in regulating the prevalence of selfish behavior, our results indicate that self-image concerns may help to keep destructive inclinations at bay as well.

Next, this paper speaks to existing experiments that compare individual and group decision-making. Prior work has made a lot of progress in studying the influence of group communication and collective decision-making in small teams (Charness and Sutter 2012; Kugler, Kausel, and Kocher 2012; Kocher, Praxmarer, and Sutter 2020). Group communication has been shown to reduce the prevalence of erroneous decisions and to lower willingness to sacrifice one's own resources to increase social welfare or to achieve fair allocation of payoffs, a pattern sometimes described as the "selfishness shift."[6] Our results provide evidence of a novel mechanism behind why decisions made in group settings are often different from individual decisions in the inter-personal domain, by documenting a systematic "destructiveness shift" caused by a perception of diffused responsibility. Further, in E.1, we empirically contrast the "selfishness shift" caused by group communication and the "destructiveness shift" caused by the perception of diffused responsibility. We show that both of these mechanisms reduce cooperation rates in the Prisoners' dilemma game, and thus are relevant for understanding the drivers of less prosocial behavior in groups. Interestingly, we show that group communication does not elevate the prevalence of destructive behavior that is costly to the decision-makers. In fact, some of our results indicate group communication and preference-aggregation may operate as a counter-acting force and help to attenuate manifestations of individual nasty inclinations in a group setting.[7]

work is that, in E.2 and E.3, we elicit preferences to be nasty on large, nationally representative samples, while existing work mostly focuses on more select groups, such as university students or individuals from disadvantaged settings.

6. Most of this literature is based on comparison of choices made by individuals and unitary teams, in which members make a joint decision. Teams have been observed to make less pro-social and honest choices in a range of tasks, including the Dictator game, the Ultimatum game, the Trust game, the PDG, and the die-rolling task—see Kugler et al. (2007), Charness and Sutter (2012), and Kocher, Schudy, and Spantig (2018) for more references. Nevertheless, choices in these tasks cannot separate the influence of own-money maximizing motivations from anti-social preferences, and thus this evidence is mute about whether group decisions affect not only selfishness but also destructiveness.

7. In terms of methods, the integrated experimental design in E.1 that aims to separately identify the effects of group context on individual behavior (Charness, Rigotti, and Rustichini 2007) and the effects of

The remainder of this paper proceeds as follows. In Section 2, we present the design and results of E.1, which studies the role of different mechanisms why a group context may increase nastiness among university students. In Sections 3 and 4, we describe the design and results of two online experiments, designed to test the robustness of the main effects among diverse samples of adults. Section 5 describes additional results, and Section 6 offers concluding thoughts.

2. Nastiness in Groups: Mechanisms (E.1)

The aim of E.1 is (i) to identify whether individuals behave more destructively toward others when they make choices in a group context than when they decide individually in isolation and (ii) to shed light on several potential mechanisms behind such an effect. Specifically, we study which features of a decision-making environment in a group setting can magnify people's tendency to engage in destructive behavior. We particularly focus on factors that can create a perception of group cover and reduced individual responsibility. We also study the effects of observability and communication within groups.

The experiment was implemented among 795 university students in Eastern Slovakia in 2019. The participants were students at the Technical University of Košice and the University of Prešov. Because there is no experimental economics laboratory in the region, the experiments were organized during regular lectures or in sessions organized at university dormitories. Around half of the students (55%) major in economics or management, while the remaining 45% major in technical disciplines, including mining, engineering, and informatics. The average age of the subjects was 21.3.[8]

2.1. Experimental Tasks

The main task we implemented is the Joy of Destruction mini-game (JDG) (Abbink and Sadrieh 2009; Abbink and Herrmann 2011), a money-burning game designed to identify destructive preferences. Two players received an endowment of €10 each, and chose whether to pay €0.5 to destroy €5 of their counterpart (destructive choice) or to keep the payoffs unchanged (non-destructive choice). The payoff matrix is presented in Panel (a) of Figure 1. The counterpart was a randomly chosen respondent who participated in the same study on a different day and was completely anonymous to the decision-maker.

Subjects made three decisions in this task. Specifically, they made choices in two conditional decisions—one in which the counterpart decided to keep the payoffs unchanged, and one when the counterpart decided to lower the decision maker's payoff.

group deliberation and preference-aggregation (Balafoutas et al. 2014; Ambrus, Greiner, and Pathak 2015) is inspired by Sutter (2009) and Feri, Irlenbusch, and Sutter (2010), who use a related design to study decision making under uncertainty and the ability to coordinate.

8. Experiment 1 was approved by the Ethical Committee at the Economics Institute of the Czech Academy of Sciences and by the Ethical Committee at the Technical University of Košice.

PANEL A: Experiment 1 (University students); in EUR

		Player B:	
		Non-destructive	Destructive
Player A:	Non-destructive	10, 10	5, 9.50
	Destructive	9.50, 5	4.5, 4.5

PANEL B: Experiment 2 (Representative sample of adult population); in EUR

		Player B:	
		Non-destructive	Destructive
Player A:	Non-destructive	2, 2	1, 1.8
	Destructive	1.8, 1	0.8, 0.8

PANEL C: Experiment 3 (Representative sample of adult population); in EUR

		Player B
		Passive
Player A:	Prosocial	1.8, 3
	Keep	2, 2
	Destructive	1.8, 1

FIGURE 1. Payoff matrices. Payoff matrices for games implemented in the experiments. The first number in the matrix cell indicates the payoff of Player A and the second number indicates the payoff of Player B.

Subjects also made an unconditional decision, that is, chose an action without knowing what their counterpart did. In the following text, we denote the choice to reduce the other's payoff in JDG as destructive.

Since a destructive choice implies reduction of both the participant's own and their counterpart's payoffs, it leads to outcomes far below the social optimum. Such behavior is typically attributed to a desire to be destructive or nasty. A crucial aspect of JDG is that the destructive choice cannot be explained by selfishness. Because the decision is a one-shot and the destructive choice is costly in that the participant has to sacrifice a part of his own payoff, the dominant strategy of a purely self-regarding player is not to engage in destructive behavior. Moreover, a destructive choice in the conditional decision, when the respondent knew that their counterpart did not act destructively, cannot be explained by inequality aversion (because the destructive choice increases inequality), negative reciprocity (because the respondent was reacting to kind rather than to unkind behavior by the counterpart), or to beliefs about the behavior of the counterpart (because the decision was conditional on the action of the counterpart).

Thus, it unambiguously reveals anti-social preferences. Therefore, in the main analysis, we primarily focus on choices in the conditional decision, when the respondent knew that their counterpart had decided not to reduce a part of respondent's payoff.

We also report results for an index that captures all choices made in JDG. It is constructed as an average of three indicator variables for whether a decision-maker acted destructively in the unconditional choice and in two conditional choices, and thus takes a value between zero and one. For completeness, we also report the results for the second conditional decision and for the unconditional decision separately. In all choices, the destructive choice is consistent with anti-social preferences, but in the conditional decision when the counterpart acted destructively, reciprocity motives can also justify destruction, and in the unconditional decision when the respondents did not know what their counterpart did, beliefs about the behavior of the counterpart may also play a role, and in combination with negative reciprocal preferences could motivate a destructive choice.[9,10]

Further, we administered the Prisoner's Dilemma game (PDG). The combination of choices in JDG and PDG helps us better understand the impact of group communication and deliberation. Previous literature has documented that groups behave less cooperatively in PDG than individuals (Wildschut et al. 2003; Charness and Sutter 2012). We are interested whether this is due to groups behaving more destructively or more selfishly, and thus whether decision-making in groups increases or reduces destructiveness levels in JDG. In PDG, two players received the same endowment of €8 and decided whether to take away €4 of their counterpart's payoff in order to increase their own payoff to €10 (defection/non-cooperative choice), or whether to keep the payoffs unchanged (cooperative choice). We again elicited two conditional decisions, an unconditional decision and beliefs. Defection is a dominant strategy for both players if they are purely self-regarding or have destructive preferences, but the socially optimal outcome is reached when both players cooperate. The payoff matrix appears in Online Appendix Figure A.1.

We used neutral framing in both tasks. For example, the question in JDG was "Do you want to pay 50 cents and reduce the other person's income by 5 euro?" In PDG, we asked "Do you want to take 4 euro from the other person to get 2 euro for yourself?" The tasks were also labeled neutrally and we never used the term "Joy of Destruction" game in the experimental protocol, because it could prime subjects to destroy. The order of JDG and PDG was randomized.[11]

9. For example, in the context of punishment experiments, Herrmann, Thoni, and Gächter (2008) show that some participants chose to pre-emptively cause financial harm to participants they expected would punish them.

10. We elicited beliefs about counterpart's behavior by asking subjects to guess which of the two strategies was selected by the counterpart. Further, in the experimental conditions in which subjects made choices in the group context (*GC_Hidden* and *GC_Observed*), we also elicited beliefs regarding the unconditional choice of the other two group members.

11. Further, we implemented a task that measured the willingness to destroy a monetary donation to participant's university, an even more extreme manifestation of destructiveness than JDG. The patterns are

2.2. Experimental Conditions

The experimental manipulations were designed to identify three effects of the decision environment on destructiveness: (i) the effect on individual choices of making choices in an anonymous group context, that is, under the cover of the group where it is unclear which group member is responsible for a group decision, (ii) the effect of observability of choices by other group members, and (iii) the effect of group decision-making, that is, of group deliberation and preference aggregation.

Subjects were randomly allocated into one of the following three conditions, using a between-subject design: *Individual*, *GroupContext_Hidden* (*GC_Hidden)*, *and GroupContext_Observed* (*GC_Observed)*. Online Appendix Table A.1 documents that the observable characteristics of participants vary little across the experimental conditions and thus randomization was successful.

In the *Individual* condition, subjects made choices in isolation and did not interact with other participants in any way. The *GC_Hidden* condition was designed to create a perception that a subject's decisions are made under the cover of the group. It contains features of the group context that are predicted to affect perception of individual responsibility for the decision, while isolating the roles of other factors. Specifically, subjects were informed that they were matched with two other individuals from a different session, and that the decision of a randomly selected member would be treated as the decision of the whole group. This can create a perception of diffused responsibility, because the subjects can exploit uncertainty about precisely who caused the nasty outcome.[12] Also, the decision was framed as a decision on behalf of the group, rather than as an individual decision, which can lower self-attribution of responsibility. To eliminate the role of social image considerations, subjects did not receive information about who the other group members were, they knew that other group members would not be informed about their choices, and that the experiment would not involve interactions with them. To avoid creating a sense of shared group identity that could affect decision-making, the group boundary was not salient: members of a group were completely anonymous to each other, were not in physical proximity, did not share any common characteristic, and could not interact with each other because they were matched across sessions and made decisions in different points in time. Consequently, we interpret the difference in behavior between *Individual* and *GC_Hidden* as reflecting the effect of group cover.

qualitatively similar to those from JDG, but the prevalence of destructive behavior in this task was very low in all treatments (0.004–0.02). Further details are provided in Online Appendix C.

12. Strictly speaking, the experimenter could figure out which group member was ultimately responsible for the group decision. However, since all the participants knew that their decisions were anonymous, and that even the experimenter did not know their name (only their experimental ID), we believe subjects were unlikely to take this into account. This is especially the case in the online experiments (E.2 and E.3), in which there was no direct contact between the experimenter and participants. Nevertheless, while the *GC_Hidden* condition was designed to substantially reduce the moral costs of acting destructively, we cannot rule out that the perception of being observed by the experimenter could still play a role and moral costs may not be eliminated completely.

The second condition in which subjects made choices in a group context, is *GC_Observed.* Subjects were randomly matched with two participants from the same session and, as in *GC_Hidden,* they were informed that a decision made by one randomly selected group member would be applied for the whole group. In contrast to *GC_Hidden*, subjects knew that their choices would be observed by the other group members and thus the difference in behavior between *GC_Observed* and *GC_Hidden* is informative about the effect of social image considerations. Group members also sat next to each other, making the group boundary more salient.

After subjects made decisions in one of the three conditions, they learned that they would make the choices again, this time jointly as a group, after having 4 minutes to discuss and reach a joint group decision (*Group* condition). Subjects were matched (*Individual*) or re-matched (*GC_Observed* and *GC_Hidden*) into groups with two other participants from the same session. Importantly, the *Group* condition came as a surprise, and thus strategic considerations arising from the expectation of group decision-making could not influence the choices made in the other conditions. At the same time, because the choices in the *Group* condition were always made last, we cannot completely rule out that the comparison of joint group decisions and decisions in the other three conditions might be influenced by order effects.[13]

The experimental counterparts were research participants from sessions which took place on other days and who made decisions in the same experimental condition as the subjects, that is, in *Individual,* the counterpart was another individual, in *GC_Hidden* and *GC_Observed*, the counterpart was a group of three participants who made decisions in the respective group context condition, and in *Group*, the counterpart was another group of three participants who made joint decisions. The payoffs per person were identical in all the conditions. At the same time, the destructive choice might have been perceived by some participants as three times more costly and more "effective" in *GroupContext* and *Group*, because it impacted three other people, than in *Individual* where it impacted one other person. To test whether this design aspect matters, in Pilot E.1, described in Online Appendix B, we implemented two sub-conditions in the *GroupContext* and *Group* conditions—one in which the counterpart was a group of three members, and one in which the counterpart was a single individual. The prevalence of destruction is very similar in both sub-conditions (Online Appendix Table A.2), and remains higher than in the *Individual* condition, documenting that a greater number of counterparts is unlikely to lead to greater destruction in *GroupContext* conditions. This pattern is in line with other work showing that people are generally insensitive to the number of counterparts who are negatively affected by their choices (Schumacher et al. 2017). Further, in E.3 described below, the

13. A way to eliminate the potential order effect would be to implement the *Group* condition using a "between-subject" design, which would, however, prevent us from studying how individual preferences are aggregated into group decisions and would require a much larger sample size. We did not randomize the order of the *Individual/GroupContext* and *Group* conditions to avoid the confounding effect of information about others' preferences obtained during the group discussion in the *Group* condition on decision-making in the *GroupContext* or *Individual* conditions.

experimental counterpart is a single individual both in *Individual* and in *GC_Hidden*, and we find the same pattern as in E.1.

2.3. Procedures

The tasks were explained using visual aids and the participants were asked four comprehension questions before making choices in JDG and in PDG. The level of understanding was high. In both tasks, 95% of subjects answered all control questions correctly, and the results are robust to excluding observations with imperfect understanding (Table 1).

The decisions were made anonymously, as the subjects submitted their choices under an experimental ID. All choices were incentivized and the subjects were paid for one randomly selected decision and for one randomly selected belief-elicitation task (€2 for a correct guess). Specific instructions for the three experimental conditions were given separately in three rooms. Online Appendix E contains the full experimental protocol.

2.4. Results

In the analysis, we first present the results for choices in the conditional decision when the respondent knew that the counterpart had decided not to act destructively, and then for the index that captures all three decisions about whether to choose the destructive option in JDG.

Effects of Group "Cover." We find that making decisions in *GC_Hidden* systematically increases the prevalence of destructive behavior (Figure 2 and Table 1) from 4% in *Individual* to 13% in *GC_Hidden* (Chi-square test, p-value < 0.001) in the conditional decision when the counterpart is not destructive. Similarly, for the index, we observe an increase in destructive behavior from 20% to 27% (Wilcoxon rank-sum test, p-value $= 0.001$). The results are robust to controlling for the order of the games, experimenter fixed effects, and a range of observable characteristics (Table 1). The effects are qualitatively similar when we focus on the unconditional decision and the second conditional decision (Online Appendix Table A.3). In line with these patterns, making decisions in a group context also matters in PDG, and leads to higher defection rates, although the effects are smaller in magnitude than in JDG (Panel B of Online Appendix Table A.3). Thus, in *GC_Hidden,* subjects become unambiguously more willing to destroy the other persons' resources, both when they pay for as well as when they gain from such actions. We refer to this empirical pattern as the "destructiveness shift."

Effects of Observability. To explore the influence of within-group observability of individual choices, we compare choices in *GC_Observed*, in which the subjects knew their group members would observe each other's decisions, and *GC_Hidden*, in which choices were made in private and subjects did not know who their group members were. We find that, for the conditional decision when the counterpart was not destructive, the

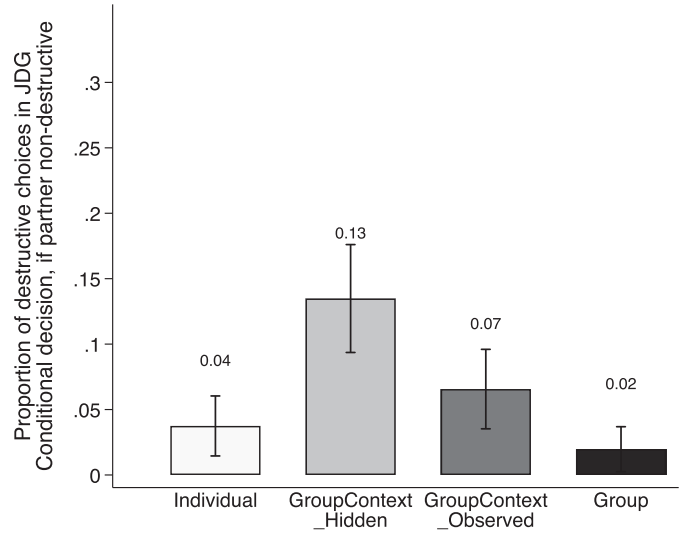
TABLE 1. Destructive behavior in groups, Experiment 1 (university students)

	Destructive behavior in the Joy of Destruction game						
Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Conditional decision, if partner non-destructive							
GroupContext_Hidden	0.10***	0.10***	0.10***	0.10***	0.05	0.10***	0.09***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)
GroupContext_Observed	0.03	0.02	0.02	0.03	0.03	0.02	0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
Group	−0.02	−0.02			−0.04**	−0.02	−0.03
	(0.01)	(0.01)			(0.02)	(0.01)	(0.02)
F-test *p*-value [GC_Hidden=GC_Observed]	0.008	0.010	0.008	0.018	0.617	0.010	0.012
F-test *p*-value [GC_Hidden=Group]	<0.001	<0.001			0.001	<0.001	<0.001
F-test *p*-value [GC_Observed=Group]	0.010	0.013			0.021	0.012	0.014
Mean Individual	0.04	0.04	0.04	0.03	0.03	0.04	0.04
Observations	1,047	1,047	790[A]	754[A]	525	1,047 (875 clusters)	1,015
Panel B: All decisions (Index: min = 0, max = 1)							
GroupContext_Hidden	0.07***	0.07***	0.07***	0.08***	0.07**	0.07***	
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	
GroupContext_Observed	0.04*	0.03	0.03	0.04	0.04	0.03	
	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.02)	
Group	−0.03*	−0.04*			−0.05*	−0.04*	
	(0.02)	(0.02)			(0.03)	(0.02)	

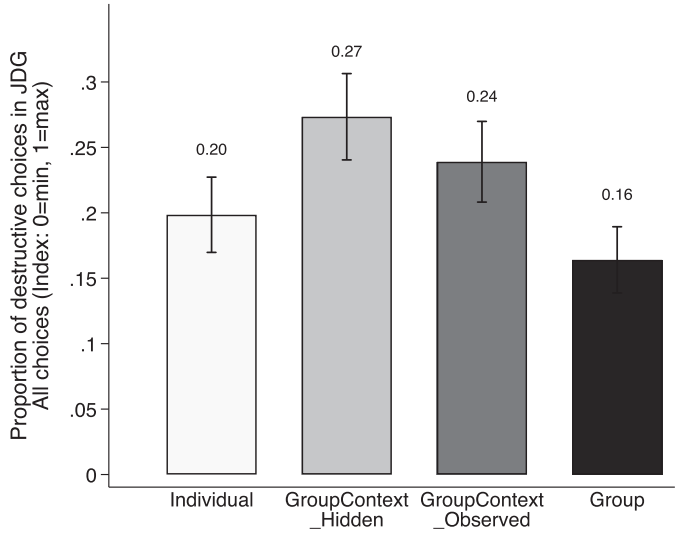
TABLE 1. (Continued)

	Destructive behavior in the Joy of Destruction game						
Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
F-test *p*-value [GC_Hidden=GC_Observed]	0.134	0.082	0.088	0.094	0.480	0.080	
F-test *p*-value [GC_Hidden=Group]	<0.001	<0.001			<0.001	<0.001	
F-test *p*-value [GC_Observed=Group]	<0.001	<0.001			0.005	<0.001	
Mean Individual	0.20	0.20	0.20	0.19	0.19	0.20	
Observations	1,046	1,046	789[A]	753[A]	525	1,046 (875 clusters)	
Specification							
Controls design		✓	✓	✓	✓	✓	✓
Controls observables			✓				
Excluding non-understanding subjects				✓			
Only JDG played first					✓		
SEs clustered at group level (GC_Observed)						✓	
Logit							✓

Notes: Effects of being in the *GroupContext_Hidden*, *GroupContext_Observed*, and *Group* conditions versus the *Individual* condition (omitted) on destructive behavior in the Joy of Destruction game. Columns (1)–(6) report OLS estimates, and Column (7) reports marginal effects from a logit. Robust standard errors in parentheses. Standard errors in Column (6) are clustered at group level for the *GroupContext_Observed* condition (three subjects per group). In Panel A, the binary dependent variable indicates destructive choice in the condition when partner is non-destructive. In Panel B, the dependent variable is the index of destructive behavior, calculated as an average of three indicator variables for whether a decision-maker acted destructively in the unconditional choice and in two conditional choices (min=0, max=1). Column (1) presents estimates without controls. Columns (2)–(7) include design-specific controls (order of games and experimenter fixed effects). Column (3) includes all observable characteristics from Online Appendix Table A.1; in case of a missing value in a dummy variable, we use another dummy variable indicating that the value is missing. Column (4) includes only subjects who answered all control questions correctly. Column (5) includes only subjects who played JDG before PDG. [A] Columns (3) and (4) do not include observations in the *Group* condition, as we cannot include the individual-level controls and measures of understanding for this condition. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.



(a) Conditional decision, if partner non-destructive



(b) All choices (Index: 0=min, 1=max)

FIGURE 2. Destructive behavior in groups, Experiment 1 (university students). Destructive behavior in the Joy of Destruction game across conditions. Panel (a) presents the conditional decision for when the partner is non-destructive, Panel (b) presents all choices (index, calculated as an average of three indicator variables for whether a decision-maker acted destructively in the unconditional choice and in two conditional choices; min=0, max=1). Bars indicate 95% confidence intervals.

prevalence of destructive behavior is 7% in *GC_Observed*, which is smaller than the prevalence of 13% in *GC_Hidden* (Chi-square test, p-value = 0.008). For the index, we also observe a drop in the prevalence of destructive behavior in *GC_Observed* as compared to *GC_Hidden*, from 27% to 24%, but the difference is not statistically significant (Wilcoxon rank-sum test, p-value = 0.180). We do not see any systematic differences in behavior in PDG across the two conditions.[14] Thus, we conclude that observability of choices does not contribute to greater destructiveness, while it may contribute to reduction of purely destructive behavior, suggesting that subjects care about their social image and are aware that destructive behavior toward non-destructive individuals in JDG is perceived by others as socially inappropriate.

Effects of Group Decision-Making. To shed light on the effects of group discussion and joint decision-making, we compare behavior in the *Group* and the *GC_Hidden* conditions. We find a clear pattern: group decision-making increases the prevalence of behavior that maximizes the payoff of the whole group (Figure 2, Table 1). In the conditional decision when the counterpart was not destructive, the prevalence of destructive behavior falls from 13% in *GC_Hidden* to 2% in *Group* (Chi-square test, p-value < 0.001). For the index, there is a reduction from 27% to 16% (Wilcoxon rank-sum test, p-value < 0.001). We observe similar effects for the unconditional choice and for the second conditional decision (Panel A of Online Appendix Table A.3). The destruction rate in *Group* is also lower than in *GC_Observed*. Thus, we find that group decision-making has the opposite, counter-acting effect from the effect of group cover on the prevalence of destructive behavior.

In PDG, the defection rate in the conditional decision when the counterpart acted cooperatively increases from 46% in *GC_Hidden* to 55% in *Group* (Chi-square test, p-value = 0.039). For the index, the prevalence of defection increases from 64% to 72% (Wilcoxon rank-sum test, p-value = 0.007). Thus, in PDG, joint decision making reinforces the effect of the group context on individual behavior—both effects contribute to greater defection of groups than of individuals. Because group decision-making decreases destructiveness in JDG and reduces cooperativeness in PDG, allowing the respondents to make decisions jointly as a group, rather than individual choices under the cover of a group, increases the likelihood of their playing strategies predicted for agents with purely self-regarding preferences, a pattern that we refer to as the "selfishness shift."

Additional Results about Preference-Aggregation in Groups. Further analysis indicates that the "selfishness shift" in *Group* is not solely an outcome of agreeing on the initial opinion of the median member, a preference-aggregation rule that

14. The finding that observability does not increase the cooperation rate in PDG, but reduces the destruction rate in JDG suggests that, while subjects care to signal that they are not anti-social, they do not find it important to signal that they are pro-social. The latter finding is consistent with other evidence, mainly focusing on lying for one's own benefit, showing that signaling pro-social behavior within groups is relatively rare (see Kocher, Schudy, and Spantig (2018) for references). These patterns are consistent with the interpretation that unambiguously nasty behavior is seen as a substantially more severe violation of social norms than is simply acting based on one's own self-interest.

mechanically lessens the prevalence of minority opinions. In Online Appendix Table A.4, we simulate a joint decision for each group and assume that it is fully determined by the choice of the median group member. Since in JDG the destructive option was chosen by a minority of subjects, the simulated decisions imply a substantial reduction in destructiveness. Interestingly, we see further reduction in destructiveness in actual group decisions than in the simulated decisions, suggesting that group deliberation and persuasion also play a role.

To shed more light on this effect, we analyze how the composition of preferences among individual group members, inferred from individual choices in *Individual*, *GC_Hidden*, and *GC_Observed*, is predictive of actual group decisions (Online Appendix Table A.5). We focus on conditional decisions when the counterpart was non-destructive, but the results are very similar for the unconditional choices. We find that 98% of the groups follow a median a priori opinion in JDG when such an opinion implies maximization of their own group payoffs, and only 2% switch to the destructive strategy after joint discussion and deliberation (Panel A of Online Appendix Table A.5). In contrast, when the initial judgement of the median group member is to destroy (there are few such cases), 89% of the groups ultimately agree on the non-destructive strategy. This pattern suggests that destructive individuals are less persuasive than non-destructive members.[15] In PDG, we also observe greater prevalence of switching to payoff-maximizing behavior after the group deliberation (Panel B of Online Appendix Table A.5). Specifically, more than twice as many groups switch to defection after group deliberation than there are groups that switch to cooperation. Thus, these results indicate that the "selfishness shift" is driven by both preference aggregation based on median opinion in groups and by the process of group deliberation.

2.5. Discussion

In this sub-section, we discuss possible mechanisms that can explain the "destructiveness shift" we observe. We argue that our results support the interpretation that the higher prevalence of destructiveness in group settings is primarily driven by a perception of diffused responsibility and cover of the group, which reduce the scruples of some participants against deviating from social norms and giving in to intrinsic urges to cause harm to others. In *GC_Hidden*, group members make decisions on behalf of an anonymous group, and more people are involved in the decision-making. These aspects of the group context are predicted to weaken subject's perceptions of their individual responsibility for the decision, compared to decision-making in *Individual*, in which it is salient that the decision-maker is fully responsible for the ultimate action. The other patterns provide further, though more indirect, support. The perception of "cover" arguably becomes weaker in environments in which group members observe

15. Because the group interactions in our study were in-person, we cannot study the deliberation process using text data, as do studies that use chat boxes to facilitate group communication (e.g. Kocher, Schudy, and Spantig 2018).

each other's choices, or when members with anti-social inclinations must persuade other group members to act in a destructive manner, so that it becomes easier to attribute anti-social inclinations to particular individuals. In line with this, we find that observability, as well as group deliberation and joint decision-making, reduces the prevalence of destructive behavior compared to behavior in *GC_Hidden*.

Next, we consider several alternative mechanisms. First, the increase in destructiveness in *GC_Hidden* is unlikely to be driven by strategic considerations related to expectations of joint decision-making in groups. The group decision-making stage came as a surprise, after subjects had completed their individual decisions in either the *Individual*, *GC_Hidden*, or *GC_Observed* condition. Further, in E.2 and E.3, described below, the joint group decision-making stage is omitted altogether, and we replicate the effects.

Second, the observed effect is also unlikely to arise due to an increase in negative reciprocity. Of the three choices made in JDG, the increase in the prevalence of destruction in *GC_Hidden*, compared to those in *Individual*, is the largest in the situation when subjects choose whether to destroy the resources of a counterpart who decided to act non-destructively, and thus when destruction cannot be explained by reciprocity and when it is particularly nasty to act in a destructive manner.[16]

Third, we consider the possibility that subjects in *GC_Hidden* may aim to please other group members by acting destructively. For this mechanism to be relevant, subjects must believe that the other group members prefer the destructive choice to the non-destructive choice. To gauge the relevance of this mechanism, we elicited subject's beliefs about the unconditional decisions made by their other two group members. In Online Appendix Figure A.2, we report the distribution of these beliefs and find that a much larger percentage of subjects (46%) in *GC_Hidden* expect both group members to be non-destructive compared to those who expect both group members to be destructive (7%). The difference is even larger in *GC_Observed*. Thus, if subjects place equal weight on the preferences of those who wanted to destroy and those who did not, a motivation to please other group members should reduce the prevalence of destructive behavior, in contrast to the pattern we observe. The observation that adding observability of behavior within groups reduces the prevalence of destructive behavior is also not consistent with this interpretation.

Fourth, when individuals are allocated to salient groups, they may develop an "in-group versus out-group" psychology (Tajfel 1981), which may motivate them to care about the relative position of their own group vis-a-vis people who belong to other groups. On one hand, compared to *Individual*, we observe an increase in the destructiveness in *GC_Observed,* a condition in which group boundaries are salient because group members personally interacted with each other. However, importantly,

16. Although the observed effect of the group "cover" does not arise due to an increase in negative reciprocity, this does not imply that reciprocal preferences do not matter in general in our experiments. In line with the literature on reciprocal preferences, we find that the likelihood of choosing the destructive option is substantially larger in the conditional decision when the counterpart is destructive, as compared to when the counterpart chose not to destroy (Online Appendix Table A.3).

the rise in destructiveness is larger in *GC_Hidden*, in which group boundary was less salient. In *GC_Hidden*, group members were anonymous to each other and there was no shared group attribute that would distinguish their group from the experimental counterpart group. In fact, the information provided about the members of their own group and their counterparts was identical—they were anonymous university students participating in the same research on a different day. Thus, this mechanism could explain our findings only if a simple reference to a "group" alone were enough to trigger in-group/out-group biases and if, at the same time, greater salience of the group boundary in *GC_Observed* compared to that in *GC_Hidden* did not further strengthen the in-group versus out-group psychology.

Finally, E.1 shows that the effects of making decisions under the cover of the group is qualitatively different from the effects of group deliberation, a mechanism identified in earlier work (Charness and Sutter 2012). Group deliberation and preference-aggregation causes a "selfishness shift": it reduces the prevalence of cooperative behavior in PDG and the prevalence of destructive behavior in JDG, making groups more prone to behave as own-money maximizers than when subjects make choices individually. In contrast, making decisions under the cover of the group causes a "destructiveness shift": individuals become generally more prone to destroy the earnings of the counterpart, both when it is privately costly (in JDG) and when they financially gain from such actions (in PDG).

3. Representative Sample of Adults (E.2)

In E.2, we focus on the main finding from E.1: when subjects make decisions under the cover of the group in the *GC_Hidden* condition rather than individually, their behavior becomes nastier. To test whether the effect holds not only for university students but also for adults across various demographic and economic groups, in E.2, we study destructiveness among a large nationally representative sample of adults in the same country.

Experiment 2 was implemented among a sample of 4,243 adults in the Slovak Republic in 2019, in cooperation with the IPSOS survey company. The sample consists of participants 18 years of age or older who were quota sampled from an actively recruited online panel to be balanced and representative of the general population based on a set of observable characteristics: age (six categories), education level (three), size of place of residence (five), and region of residence (eight). As compared to the census statistics, in our sample males are slightly under-represented (45%), as are respondents with lower education levels, who are more difficult to reach in an online survey (Online Appendix Table A.6). Distributions of other characteristics are similar.[17]

17. Experiment 2 was approved by the Ethical Committee at the Economics Institute of the Czech Academy of Sciences and by the Ethical Committee at the Technical University of Košice.

3.1. Experimental Tasks

As in E.1, the subjects participated in the Joy of Destruction mini-game (JDG) and in the PDG in randomized order. In JDG, both players received €2 and made a decision whether to pay €0.2 to destroy €1 belonging to the counterpart, or whether to keep the payoffs unchanged (Panel (b) of Figure 1). In PDG, both players received €1.6 and decided whether to take €0.8 of the counterpart to gain €0.4 for themselves, or to keep the payoffs unchanged.

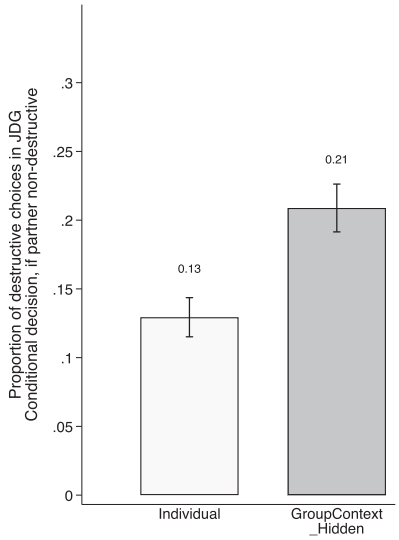
In each game, participants made two conditional choices, stating what they wanted to do in case the counterpart reduced their payoff and in case the counterpart kept the payoffs unchanged, and an unconditional decision (as in E.1). All choices were incentivized. The participants were paid for one randomly selected decision.

3.2. Experimental Conditions

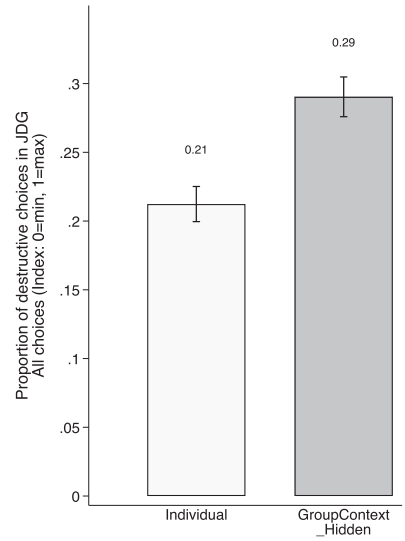
In E.2, we implemented only the *Individual* and *GC_Hidden* conditions, using a between-subject design.[18] The *GC_Hidden* condition is very similar to the *GC_Hidden* condition implemented in E.1. Subjects made choices on behalf of a group of three individuals. They knew that the other two group members also participated in the survey, but they were completely anonymous, could be from anywhere in the country, and the subjects could not interact with them in any way. The participants were informed that the decision of one randomly selected group member would be relevant for the whole group. In *Individual*, the counterpart was another individual, while in *GC_Hidden* the counterpart was another group of three participants who also made decisions in a group context. Online Appendix E contains the full experimental protocol.

Because we have such a diverse sample in E.2, we gathered a rich set of observable participant characteristics. In Online Appendix Table A.6, we find no systematic differences in observable characteristics across *Individual* and *GC_Hidden*. Nevertheless, due to a misunderstanding in our communication with the survey company, the subjects were not allocated purely randomly to these two conditions. For each of the conditions, the survey company followed a separate detailed quota sampling procedure to generate comparable groups. Thus, although we do not detect statistically significant differences for any of the 33 observable characteristics, we cannot fully rule out that there might be differences in some unobservable characteristics. At the same time, we find it reassuring that the estimated coefficients are virtually intact when we control for various sets of observable characteristics, the estimated effects hold across many sub-groups, and they are qualitatively similar to the patterns observed in E.1 and E.3.

18. This design choice was guided by the following considerations. First, E.1 showed that the "destructiveness shift" in groups is driven by contextual features that create perceptions of diffused responsibility embedded in *CG_Hidden*. Another aspect was feasibility; it was not possible to implement the *GC_Observed* and *Group* conditions using an online survey infrastructure, because they require face-to-face, real-time interactions between group members.



(a) Conditional decision, if partner non-destructive



(b) All choices (Index: 0=min, 1=max)

FIGURE 3. Destructive behavior in groups, Experiment 2 (representative sample of adult population). Destructive behavior in the Joy of Destruction game across conditions. Panel (a) presents the conditional decision for when the partner is non-destructive, Panel (b) presents all choices (index, calculated as an average of three indicator variables for whether a decision-maker acted destructively in the unconditional choice and in two conditional choices; min=0, max=1). Bars indicate 95% confidence intervals.

3.3. Results

We find that making a decision in *GC_Hidden* has an important influence on behavior (Figure 3 and Table 2), replicating the findings of E.1. We first analyze conditional decisions when the counterpart was not destructive. In JDG (Panel A of Online Appendix Table A.7), participants in *Individual* chose the destructive option in 13% of cases, while participants in *GC_Hidden* chose the destructive option in 21% of cases. Thus, making a decision under the cover of the group rather than individually increases the prevalence of destructive behavior by around 62% (Chi-square test, p-value < 0.001, Figure 3 and Online Appendix Table A.7). We observe similar effects for the index of destructive behavior, constructed as the average of three indicator variables of a destructive choice (two conditional choices and unconditional choice), as well as for all three choices from which the index is constructed. For the index, the prevalence of destructive behavior increases from 21% in *Individual* to 29% in *GC_Hidden* (Wilcoxon rank-sum test, p-value < 0.001, Figure 3 and Online Appendix Table A.7). Because the observed effect is largest when subjects chose whether to destroy the resources of a counterpart who decided to act non-destructively, as in E.1, these patterns show that the greater prevalence of

TABLE 2. Destructive behavior in groups, Experiment 2 (representative sample of adult population).

	Destructive behavior in the Joy of Destruction game				
Dependent variable	(1)	(2)	(3)	(4)	(5)
Panel A: Conditional decision, if partner non-destructive					
GroupContext_Hidden	0.08***	0.08***	0.07***	0.08***	0.08***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Mean Individual	0.13	0.13	0.13	0.11	0.13
Observations	4,243	4,243	4,243	2,067	4,243
Panel B: All decisions (Index: 0=min, 1=max)					
GroupContext_Hidden	0.08***	0.08***	0.07***	0.08***	
	(0.01)	(0.01)	(0.01)	(0.01)	
Mean Individual	0.21	0.21	0.21	0.18	
Observations	4,243	4,243	4,243	2,067	
Specification					
Controls design		✓	✓	✓	✓
Controls observables			✓		
Only JDG played first				✓	
Logit					✓

Notes: Effects of being in the *GroupContext_Hidden* condition versus the *Individual* condition (omitted) on destructive behavior in the Joy of Destruction game. Columns (1)–(4) report OLS estimates, and Column (5) reports marginal effects from a logit. Robust standard errors in parentheses. In Panel A, the binary dependent variable indicates destructive choice in the condition when partner is non-destructive. In Panel B, the dependent variable is the index of destructive behavior, calculated as an average of three indicator variables for whether a decision-maker acted destructively in the unconditional choice and in two conditional choices (min=0, max=1). Column (1) presents estimates without controls. Columns (2)–(5) include design-specific controls (order of games). Column (3) includes all observable characteristics from Online Appendix Table A.6, controlling for gender, age group, education, municipality size, region, and household income. Column (4) includes only subjects who played JDG before PDG. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

destructiveness in the *GC_Hidden* cannot be explained by a stronger preference to retaliate against harmful behavior by the counterpart, or by differences in beliefs about the behavior of the counterpart, but rather are explained by reduced scruples against deviating from social norms and giving in to an intrinsic preference to do harm. The results are robust to controlling for the order of the games and hold when we control for a range of observable characteristics (Table 2).

Making decisions in *GC_Hidden* also influences behavior in PDG (Panel B of Online Appendix Table A.7). For conditional decisions when the counterpart acted cooperatively, the prevalence of defection increases from 23% in *Individual* to 34% in *GC_Hidden* (Chi-square test, p-value < 0.001). When we focus on the index, the defection rate rises from 35% to 42% (Wilcoxon rank-sum test, p-value < 0.001). Similar effects are observed for each of the three decisions.

In E.2, we implemented an additional task to address the concern that the observed differences in the prevalence of destructive behavior in JDG (or defection in PDG) could be driven by a stronger urge to be active in a group setting in *GC_Hidden*,

by choosing the destructive option instead of the default allocation. In the "action-bias" task, the participants could choose between a status quo and an active choice, but the task did not have payoff consequences for the experimental counterpart, and thus destructive preferences should not affect decision-making. At the end of the survey, the respondents were asked to select one of three emoticons reflecting how they liked the survey. The emoticon by default was either yellow or blue. The respondents could choose to keep the color (opt for a passive option) or to change it (an active option). We find that in the "action-bias" task, participants in *GC_Hidden* are not more likely to change the color of an emoticon than are those allocated to *Individual* (Online Appendix Table A.8),[19] indicating that the effects of *GC_Hidden* are unlikely to be driven by a greater tendency to choose an active option than to passively stick with the default.

Finally, we take advantage of the size and breadth of our sample and perform a sub-sample analysis based on observable characteristics, testing whether the observed "destructiveness shift" is a phenomenon that is characteristic of a certain well-defined group, or whether it reflects a behavioral response that is generalizable across demographics. Note that, given the size of the sample, we can detect meaningful treatment effects even within each of the groups. In terms of age, we divide the sample into six groups: 18–24, 25–34, 35–44, 45–54, 55–64, and 65 and up. In terms of education, we distinguish three groups based on the highest completed level of education: primary or lower secondary, upper secondary, and tertiary. In terms of income, we divide the sample into quartiles. In terms of the size of municipality where respondents live, we distinguish five groups based on the number of inhabitants in thousands: less than 1, 1–5, 5–20, 20–100, and more than 100. Finally, we classify participants based on their political party preferences across two dimensions: economic right/left wing (right, middle, and left) and liberal/conservative (liberal, neutral, and conservative).[20]

The effects are robust. We find that making decisions in *GC_Hidden* increases destructive behavior among all the demographic sub-groups we study (Figure 4 and Online Appendix Table A.9). For the conditional decision when the counterpart was not destructive, the effects are positive and statistically significant for 23 out of 26 sub-groups (except subjects who are 45–54 years old, 65 or older, and those from the 4th income quartile). For the index, the effects are positive and statistically significant for all subgroups except one (Online Appendix Figure A.3). The effects on behavior in PDG are similarly robust (Online Appendix Figure A.4).

19. We arrive at the same conclusions in E.1, where we measured the "action bias" using a different task in the *Individual* and *Group* conditions. Online Appendix C provides more details.

20. We provide more details about the classification of political parties in Online Appendix Table A.9.

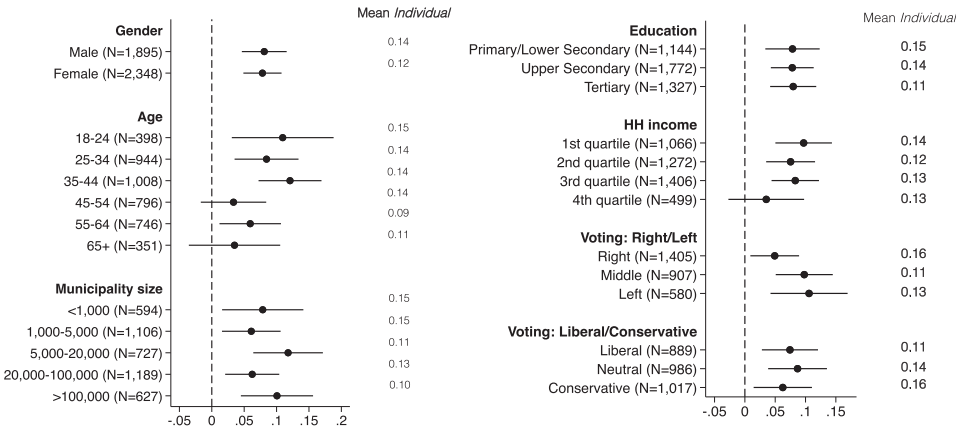


FIGURE 4. Effect of a group context on destructive behavior in the Joy of Destruction game: Heterogeneity across sub-groups, Experiment 2 (representative sample of adult population). Estimated effect of being in the *GroupContext_Hidden* condition, compared to the *Individual* condition on destructive behavior, across different demographic sub-groups (separate OLS regressions by sub-groups). Horizontal bars indicate 95% confidence intervals. The dependent variable is the destructive choice in the Joy of Destruction game, for the condition when the partner is non-destructive. The corresponding regression table is provided in the Online Appendix (Table A.9).

4. Robustness: Task with Options to Destroy and to Give (E.3)

In the previous experiments, we estimated the effects of a group context on nasty behavior in the JDG. A potential concern is that the decision situation in this task places conflict and the option to destroy at the top of the respondent's minds, an aspect that may not be present in many real-life situations. Though neutral language is used, the asymmetric choice space where the respondents can choose either to cause harm or to be passive, but not to act kindly, may signal that the task is related to measuring destructive behavior. Further, the experimental counterpart, who may choose to act maliciously toward a decision-maker, could be perceived as an adversary. The perception of a conflictual nature of the interaction may remain present to some extent even in the conditional decision (our primary measure of nasty behavior in E.1 and E.2), in which the participant knows that the counterpart did not choose to destroy, but could have done so. These features of the JDG may affect the choices of the respondents, specifically by leading to a relatively high proportion of destructive choices, and they could potentially also affect the treatment effect, if people are more sensitive to cues of conflict in the group context compared to when they act individually.

Therefore, in E.3, we adapted the task so that the decision situation is unlikely to prime conflict. We added a third option to make the decision situation symmetric—the respondents could choose to keep the payoffs as they were, to reduce the counterpart's payoff at their own expense, or to increase the counterpart's payoff at their own

expense. Further, the counterpart was completely passive; s/he could not cause any harm to the respondents, and thus could not be perceived as an adversary.

4.1. Experimental Design

Experiment 3 was implemented among a sample of 3,349 adults in the Slovak Republic in 2022. The data collection was conducted online in cooperation with the AKO survey company. The sample consists of participants 18 years of age or older who were quota sampled to obtain a sample that would be representative of the general population in terms of gender, age (five categories), education level (three), size of place of residence (five), and region of residence (eight). Indeed, the distributions of these characteristics in our sample are similar to 2021 census statistics (Online Appendix Table A.10).[21]

The subjects made a choice in a single task. There were two players who received €2 each. The respondents made a decision whether to (i) pay €0.20 to destroy €1 belonging to the counterpart and thus reduce the counterpart's payoff to €1, (ii) keep the payoff unchanged, or (iii) pay €0.2 to increase the counterpart's payoff by €1 to €3 (Panel (c) of Figure 1). Thus, the stakes were identical to those in E.2, and the task allows us to identify unambiguously nasty behavior, as in E.1 and E.2, that cannot be explained by selfishness, inequality aversion, and reciprocity.[22] Furthermore, the counterpart did not participate in the same study and was completely passive, that is, did not make any choice, so the role of beliefs about the counterpart's behavior was closed by design. The choices were incentivized: There was a 50% probability that the decision of a given individual or group of three members would be implemented.

The subjects were allocated either to the *Individual* or the *GC_Hidden* condition, as in E.2, using a between-subject design.[23] Both in *Individual* and in *GC_Hidden,* the counterpart was another individual. Online Appendix E contains the full experimental protocol. The randomization worked well, which is documented by the fact that

21. Experiment 3 was approved by the Ethical Committee at the Technical University of Košice and by the Ethical Committee at the School of Education, Durham University.

22. We also implemented a second version of the task in which the choices to change the counterpart's payoff came at no cost to the respondents. Specifically, the respondent could choose to keep the payoff of their counterpart at EUR 2, to increase it to EUR 3, or to reduce it to EUR 1, always keeping EUR 2 for him/herself. This version of the task was implemented among a randomly selected half of the sample. In the analysis, we are primarily interested in choices in the costly version, because the set-up is the most similar to the decision situation in E.1 and E.2, and allows us to identify unambiguously nasty behavior. In contrast, the choice to destroy in the costless version is consistent not only with nasty behavior, but also with selfish preferences. We implemented the costless version because a supplementary data collection in which we test whether providing three versus two options matters (which we mention later in the text) indicated that relatively few people may be willing to incur a cost in order to reduce the payoff of a counterpart who is completely passive. In the end, we are sufficiently powered to detect effects of the group context in the costly version, and so in the main analysis we focus on this cleaner measure of nasty behavior. We arrive at similar conclusions when we pool choices from both versions of the task, and the effects of the group context are qualitatively similar (though not statistically significant) when estimated separately in the costless version (see Table 3 and Online Appendix Table A.11).

23. The *Individual/GC_Hidden* conditions and costly/costless versions of the task were manipulated orthogonally in a 2x2 between-subject design.

we do not find systematic differences across conditions for a rich set of observable characteristics (Online Appendix Table A.10). There is some imbalance regarding the region of origin; the results reported below are robust when controlling for all observable characteristics.

4.2. Main Results

We find that the group context has systematic effects on allocations that are in line with the findings from E.1 and E.2. We first analyze the effects on the amount allocated to the counterpart and then explore whether the differences in allocations are driven by a lower preference to give or a greater preference to destroy. In the main analysis, we focus on choices in which participants must sacrifice their own money to reduce the earnings of their counterpart, and thus a decision to do so provides a clean measure of a preference to be nasty.

In Table 3, Panel B, we show that participants in *CG_Hidden* allocate systematically less money to a counterpart than do participants in *Individual* (OLS, p-value $= 0.030$). The results are robust to various specifications.

Next, importantly, we show that the differences in allocations are driven by both a greater tendency to engage in nasty behavior and by a somewhat lower willingness to give to others. Specifically, in *Individual,* 3.3% of subjects chose to destroy earnings of the counterpart, and the prevalence of destruction increases by 2 percentage points to 5.2% in *GC_Hidden* (Chi-square test, p-value $= 0.050$, Figure 5 and Online Appendix Table A.11), a 58% increase in the prevalence of destructive behavior. This difference is robust to adding control variables and excluding subjects who did not pass all attention checks (Panel A of Table 3). We also observe some reduction in the prevalence of helping behavior, from 14.7% in *Individual* to 12.3% in the *GC_Hidden*, although this difference is not statistically significant at conventional levels (Chi-square test, p-value $= 0.147$, Online Appendix Table A.11).[24]

Overall, the results show that the behavioral regularity identified in E.1 and E.2 is not specific to the situation in which the decision-maker can only either harm or be passive. Together, the findings from E.1–E.3 suggest that the effect of perceived group "cover" on greater willingness to engage in destructive behavior holds in diverse situations—when the decision-maker can only harm, when she/he can harm or help, when the counterpart is active and can be seen as an adversary, and when the counterpart is completely passive.

Although the main goal of E.3 is to test the robustness of the causal effects of perceived group "cover" in the modified task, we also discuss the observed levels of destructive behavior. We find that the prevalence of nasty behavior in E.3 is generally much lower than in E.2, even though both samples were quota sampled

24. Online Appendix Figure A.5 reports the estimated effects for various sub-groups. Although the point estimates do not suggest a systematic heterogeneity in effects across various observable participant characteristics, we note that most of the coefficients for individual sub-groups are not statistically significant at conventional levels due to reduced sample size.

TABLE 3. Destructive behavior in groups, Experiment 3 (representative sample of adult population).

	Behavior in the Joy of Destruction game with a prosocial option					
Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Destructive choice						
GroupContext_Hidden	0.019*	0.019*	0.019*	0.016*	0.014*	0.007
	(0.010)	(0.010)	(0.011)	(0.008)	(0.007)	(0.011)
Mean Individual	0.033	0.033	0.028	0.033	0.033	0.051
Observations	1,689	1,689	1,330	1,689	3,349	1,660
Panel B: Allocation to Player B						
GroupContext_Hidden	−0.043**	−0.042**	−0.031		−0.031**	−0.017
	(0.020)	(0.020)	(0.023)		(0.015)	(0.024)
Mean Individual	0.11	0.11	0.12		0.11	0.17
Observations	1,689	1,689	1,330		3,349	1,660
Specification						
Design: standard (costly destruction)	√	√	√	√	√	
Design: costless destruction					√	√
Controls observables		√	√	√	√	√
Excluding non-attentive subjects			√			
Logit				√		

Notes: Effects of being in the *GroupContext_Hidden* condition vs. the *Individual* condition (omitted) on behavior in the Joy of Destruction game. Columns (1)–(3) and (5)–(6) report OLS estimates, and Column (4) reports marginal effects from a logit. Robust standard errors in parentheses. In Panel A, the binary dependent variable indicates destructive choice in the game. In Panel B, the dependent variable is allocation to Player B, equal to 0 if the decision maker kept the reward unchanged, −1 if he made a destructive choice and 1 if he made a prosocial choice. Columns (1)–(4) show results for the costly version of the task, which was the standard in Experiment 1 and 2. Column (5) pools observations from the costly and costless version of the task, controlling for a binary variable indicating that a costless version was played. Column (6) shows results for the costless version of the task. Column (1) presents estimates without controls. Columns (2)–(6) includes all observable characteristics from Online Appendix Table A.10, controlling for gender, age group, education, municipality size, region, and household income. Column (3) includes only subjects who passed both attention checks. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

to be representative of the Slovak adult population in terms of numerous observable characteristics. In the *Individual* condition in E.2, the likelihood of choosing the destructive option is 13% in the conditional decision, when decision-makers knew that their counterpart chose not to be destructive, whereas the likelihood of choosing the destructive option is only 3.3% in E.3. This large difference is consistent with the interpretation that the asymmetric choice space in the standard JDG may create experimenter demand effects or prime subjects to think about conflict, and thus may inflate the measured levels of destructiveness.[25] This would suggest that the levels of destructiveness observed in the standard JDG, in which subjects can only harm, should

25. To explore this more directly, in 2022, we implemented a supplementary experiment ($N = 808$), described in more detail in Online Appendix D, in which we tested whether simply adding an option to increase the counterpart's payoff in JDG reduces destructive behavior in the *Individual* condition. Two

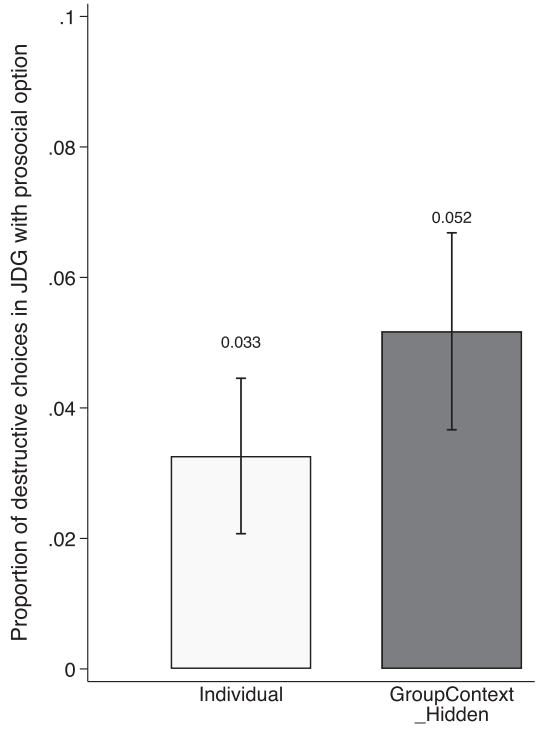


FIGURE 5. Destructive behavior in groups, Experiment 3 (representative sample of adult population). Destructive behavior in the Joy of Destruction game with the prosocial option across conditions. Bars indicate 95% confidence intervals. The Figure displays the proportion of destructive choices in the costly version of the task. The results for the proportion of prosocial choices and for the costless version of the task are available in Online Appendix Table A.11.

be interpreted with caution, and that using the version with expanded choice space might provide a more accurate picture of the prevalence of anti-social preferences.

Nevertheless, we cannot rule out other explanations. The counterpart was passive E.3, while in E.2 s/he was active and consequently subjects could have pictured the counterpart as an adversary, rather than as a neutral person. This interpretation

players received EUR 2 each. In Variant A ($N = 405$), they had only two options, as in standard JDG: (i) not to destroy or (ii) to destroy, by paying EUR 0.20 in order to reduce the counterpart's reward by EUR 1. In Variant B ($N = 403$), there was a third option, where subjects could pay EUR 0.2 to increase their counterpart's payoff by EUR 1 to EUR 3. The counterpart was passive and thus Variant B was identical to the task that we later implemented in E.3. All subjects made choices only in the *Individual* condition. Overall, the evidence is not conclusive. On one hand, the prevalence of destructive choices is 4% in Variant A and 3.2% in Variant B and thus, qualitatively, this pattern is consistent with the restricted choice space making people more prone to destroy. On the other hand, given the generally low levels of destruction in this sample (and thus not much variation in the outcome variable) and smaller sample size, we are underpowered to detect meaningful effect sizes with sufficient confidence and this difference is far from being statistically significant (Chi-square test, p-value = 0.580).

is consistent with recent work showing that competitive settings make people less prone to identify with each other than cooperative settings (Hagenbach, Kranton, and Lee 2023). In addition, we cannot rule out that the differences in the levels of destructiveness across E.2 and E.3 are simply due to differences in samples and timing.

5. Evidence from Pilot Experiments Among Adolescents

Here, we briefly summarize the design and main patterns of two pilot experiments among adolescents in Slovakia and Uganda, which were implemented before E.1–E.3 and which motivated this research agenda. Details of the design of these experiments and a complete discussion of the findings appear in Online Appendix B.

The pilot experiments focus on destructive behavior of adolescents from disadvantaged regions. We consider this subject pool a natural starting point for this enquiry, as crime, urban riots, and other forms of destructive behavior are particularly common among youth with lower socioeconomic status (Deming 2011). The experiments were conducted in 2013 in schools in Eastern Slovakia (PE.1, $N = 630$) and in rural Uganda (PE.2, $N = 1{,}679$). As in E.1 and E.2, subjects made choices in JDG and PDG.

To elicit behavior under a group "cover," we employed a somewhat different approach than in the main experiments. We asked subjects to privately state their preferences about how they wanted their group of three individuals to decide when making a joint decision in JDG and PDG, prior to discussing a joint decision with other group members.[26] In both settings, the destructive (in JDG) and non-cooperative (in PDG) options are more prevalent in this type of group context than when choices are made individually on the subject's own behalf (Online Appendix Tables A.12 and A.13).

Therefore, the findings are consistent with the observed "destructiveness shift" documented in E.1–E.3, and suggest that the main effect also holds among adolescents and is robust across two different cultures (Slovakia and Uganda). Nevertheless, the experimental design used in the pilot experiments cannot rule out the role of several alternative mechanisms other than the role of the perceived group cover, such as the effects of expected communication inside groups or social image considerations—limitations which are not present in the three main experiments.

It is also noteworthy that the prevalence of destructive behavior is generally higher among Slovak and Ugandan samples of adolescents (Online Appendix Tables A.12 and A.13) than in all three samples of adults in Slovakia (Online Appendix Tables A.3, A.7, and A.11). In *Individual*, the prevalence of destructive behavior conditional on the counterpart not being destructive is 4% among Slovak university students (E.1), 13% among Slovak adults (E.2), and increases to 29% among Slovak adolescents (PE.1) and 44% among Ugandan adolescents (PE.2). Even though neither sample of adolescents is

26. In addition, in the experiment with Slovak adolescents (PE.1), we elicited incentivized decisions in a condition similar to *GC_Observed*. We arrive at similar conclusions using this measure.

nationally representative for this age group (they are all from relatively disadvantaged areas), the observed difference in destructive behavior in the same task implemented in the same country (Slovakia) indicates that adolescence might be a developmental period in which antisocial preferences or norms (or the desire to act against social norms) might be relatively common, and that these inclinations decline in early adulthood. This pattern is broadly in line with earlier work documenting that efficiency concerns in allocation tasks rise substantially with age during late adolescence (Sutter, Zoller, and Glätzle-Rützler 2019).

More generally, in light of extensive evidence suggesting that perceptions of social norms strongly shape social behavior (e.g. Fehr and Gächter 2002; Krupka and Weber 2013), an interesting open question for future research is to explore whether norms regulating anti-social behavior vary systematically across age groups. Because we have not elicited perceptions of social norms in E.1 and E.2, we cannot test whether there are differences in the perceptions of and the desire to adhere to norms that could explain the higher observed prevalence of destructive behavior among adolescents compared to adults. Only in E.3., we used a simplified version of the approach developed by Krupka and Weber (2013) to elicit perceptions of norms for behavior in the allocation task. We asked participants to guess the views of ten individuals who did not make the allocation decisions themselves, in terms of their appropriateness ratings of choosing the destructive option. As expected, we find that most participants (72%) believe that a majority of people perceive the choice of the destructive option as being socially inappropriate, and that appropriateness ratings intuitively correlate with choices in the allocation task (Spearman's rho between a destructive choice and its appropriateness rating $= 0.15$, p-value < 0.001).

6. Conclusions

This paper provides evidence that people are more prone to engage in nasty behavior, causing harm to other people at their own expense, when they make decisions in a group context, in which it is blurred who is responsible for the harmful action, as compared to when making their own individual decisions. We establish this behavioral regularity across a broad range of demographic and socio-economic groups, by conducting a series of large-scale experiments among university students, adolescents, and nationally representative samples of adults—more than ten thousand subjects in total.

The findings have several implications. First, they provide empirical support for the idea that the perception of individual responsibility is an important regulator of the dark side of human personality. They illuminate that, in a group context, a decision environment that is ubiquitous in the real world, the perception of individual responsibility can easily be diluted, leading to socially undesirable impacts on behavior. Second, one of our experiments suggests that the greater risk of nasty behavior in group settings is particularly relevant for social structures such as crowds or mobs where there is little space for observability of the actions of individual members or for deliberations by the group deliberation, by showing that observability and group

communication may attenuate the prevalence of nasty behavior in group settings. Third, in terms of practical implications, our findings suggest that organizations and public bodies seeking to limit obstructionism and other manifestations of nasty inclinations may want to create environments that foster perceptions of individual responsibility. This can be done by, for example, framing decisions as the choices of specific individuals, rather than presenting them as decisions of whole units, teams or committees, and by making the ways the actions of individual group members translate into group outcomes more transparent.

Finally, inspired by evidence showing that people tend to seek out environments in which they feel less pressure to behave according to social norms (e.g. DellaVigna, List, and Malmendier 2012), we end with speculation about a potential endogenous response to the group "cover" effect we identify. If people are aware that, in certain environments, they feel less constrained by self-image concerns, they may rationally choose to join (or create) such an environment. Thus, for example, some people may join gangs or public protests not necessarily because they identify with them, but because the group or crowd environment provides more anonymity and ease to act upon their nasty inclinations without taxing their self-image. We suspect that this could be a fruitful avenue to explore.

References

Abbink, Klaus and Benedikt Herrmann (2011). "The Moral Costs of Nastiness." *Economic Inquiry*, 49, 631–633.

Abbink, Klaus and Abdolkarim Sadrieh (2009). "The Pleasure of Being Nasty." *Economics Letters*, 105, 306–308.

Allport, Gordon W. (1968). "The Historical Background of Modern Social Psychology." In *The Handbook of Social Psychology*, edited by G. Lindzey and E. Aronson, Addison-Wesley, Reading, MA, pp. 1–80.

Almas, Ingvild, Alexander Cappelen, and Bertil Tungodden (2020). "Cutthroat Capitalism versus Cuddly Socialism: Are Americans More Meritocratic and Efficiency-Seeking than Scandinavians?" *Journal of Political Economy*, 128, 1753–1788.

Ambrus, Attila, Ben Greiner, and Parag A. Pathak (2015). "How Individual Preferences Are Aggregated in Groups: An Experimental Study." *Journal of Public Economics,* 129, 1–13.

Andreoni, James and B. Douglas Bernheim (2009). "Social Image and the 50–50 Norm: A Theoretical and Experimental Analysis of Audience Effects." *Econometrica,* 77, 1607–36.

Balafoutas, Loukas, Rudolf Kerschbamer, Martin Kocher, and Matthias Sutter (2014). "Revealed Distributional Preferences: Individuals vs. Teams." *Journal of Economic Behavior & Organization,* 108, 319–330.

Publius (1948). "Federalist No. 55." In *The Federalist*, edited by M. Beloff. Blackwell, Oxford, pp. 283–287. (Original work published 1788).

Benabou, Roland, Armin Falk, and Jean Tirole (2018). "Narratives, Imperatives, and Moral Reasoning." NBER Working Paper No. 24798.

Benabou, Roland and Jean Tirole (2011). "Identity, Moral and Taboos: Beliefs as Assets." *The Quarterly Journal of Economics,* 126, 805–855.

Charness, Gary, Edi Karni, and Dan Levin (2010). "On the Conjunction Fallacy in Probability Judgment: New Experimental Evidence Regarding Linda." *Games and Economic Behavior*, 68, 551–6.

Charness, Gary, Luca Rigotti, and Aldo Rustichini (2007). "Individual Behavior and Group Membership." *American Economic Review*, 97(4), 1340–52.

Charness, Gary and Matthias Sutter (2012). "Groups Make Better Self-Interested Decisions." *Journal of Economic Perspectives,* 26(3), 157–176.
Dana, Jason, George Loewenstein, and Roberto Weber (2011). "Ethical Immunity: How People Violate Their Own Moral Standards Without Feeling They Are Doing So." In *Behavioral Ethics: Ideas on an Emerging Field*, edited by Ann E., Tenbrunsel, and David De Cremer. Routledge, New York, pp. 197–216.
Dana, Jason, Roberto Weber, and Jason Xi Kuang (2006). "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness." *Economic Theory*, 33, 67–80.
Darley, John M. and Bibb Latane (1968). "Bystander Intervention in Emergencies: Diffusion of Responsibility." *Journal of Personality and Social Psychology*, 8, 377–383.
DellaVigna, Stefano, John A. List, and Ulrike Malmendier (2012). "Testing for Altruism and Social Pressure in Charitable Giving." *The Quarterly Journal of Economics*, 127, 1–56.
Deming, David J. (2011). "Better Schools, Less Crime?" *The Quarterly Journal of Economics*, 126, 2063–2115.
Falk, Armin (2021). "Facing Yourself : A Note on Self-Image." *Journal of Economic Behavior & Organization*, 186, 724–734.
Fehr, Ernst and Simon Gächter (2002). "Altruistic Punishment in Humans." *Nature*, 415, 137–140.
Fehr, Ernst, Karla Hoff, and Mayursh Kshetramade (2008). "Spite and Development." *American Economic Review*, 98(2), 494–9.
Feri, Francesco, Bernd Irlenbusch, and Matthias Sutter (2010). "Efficiency Gains from Team-Based Coordination—Large-Scale Experimental Evidence." *American Economic Review*, 100(4), 1892–1912.
Gangadharan, Lata, Philip J. Grossman, and Joe Vecci (2020). "Antisocial Behavior in the Workplace." In *Handbook of Labor, Human Resources and Population Economics*, edited by Klaus F. Zimmermann. Springer, pp. 1–26.
Hagenbach, Jeanne, Rachel Kranton, and Victoria Lee (2022). "Competition, Cooperation, and Social Perceptions." hal-03792554v1
Herrmann, Benedikt, Christian Thoni, and Simon Gächter (2008). "Antisocial Punishment across Societies." *Science,* 319, 1362–1367.
Hershcovis, M. Sandy, Nick Turner, Julian Barling, Kara A. Arnold, Kathryne E. Dupré, Michelle Inness, Manon Mireille Leblanc, and Miro Sivanathan (2007). "Predicting Workplace Aggression : A Meta-Analysis." *Journal of Applied Psychology,* 92, 228–38.
Karakostas, Alexandros, Nhu Tran, and Daniel John Zizzo (2022). "Experimental Insights on Anti-Social Behavior: Two Meta-Analyses." *Discussion Papers No.* 658, School of Economics, University of Queensland, Australia.
Kocher, Martin G., Matthias Praxmarer, and Matthias Sutter (2020). "Team Decision-Making." In *Handbook of Labor, Human Resources and Population Economics*, edited by K.F. Zimmermann. Springer, Heidelberg, pp. 1–25.
Kocher, Martin G., Simeon Schudy, and Lisa Spantig (2018). "I Lie? We Lie! Why? Experimental Evidence on a Dishonesty Shift in Groups." *Management Science,* 64(9), 3995–4008.
Krupka, Erin L. and Roberto Weber (2013). "Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?" *Journal of the European Economic Association*, 11, 495–524.
Kugler, Tamar, Gary Bornstein, Martin G. Kocher, and Matthias Sutter (2007). "Trust between Individuals and Groups: Groups Are Less Trusting than Individuals but Just as Trustworthy." *Journal of Economic Psychology*, 28, 646–657.
Kugler, Tamar, Edgar E. Kausel, and Martin G. Kocher (2012). "Are Groups More Rational than Individuals? A Review of Interactive Decision Making in Groups." *Cognitive Science,* 3, 471–482.
Le Bon, Gustave (1895). *The Crowd*. New York, NY: Viking, New York, NY.
Neuman, Joel H. and Robert A. Baron (1998). "Workplace Violence and Workplace Aggression: Evidence Concerning Specific Forms, Potential Causes, and Preferred Targets." *Journal of Management,* 24, 391–419.

Prediger, Sebastian, Bjorn Vollan, and Benedikt Herrman (2014). "Resource Scarcity and Antisocial Behavior." *Journal of Public Economics,* 119, 1–9.
Sanjaya, Muhammad Ryan (2023). "Antisocial Behavior in Experiments: What Have We Learned from the Past Two Decades?" *Research in Economics,* 77, 104–115.
Schumacher, Heiner, Iris Kesternich, Michael Kosfeld, and Joachim Winter (2017). "One, Two, Many-Insensitivity to Group Size in Games with Concentrated Benefits and Dispersed Costs." *Review of Economic Studies*, 84, 1346–77.
Sutter, Matthias (2009). "Individual Behavior and Group Membership: Comment." *American Economic Review*, 99(5), 2247–2257.
Sutter, Matthias, Claudia Zoller, and Daniela Glätzle-Rützler (2019). "Economic Behavior of Children and Adolescents—A First Survey of Experimental Economics Results." *European Economic Review*, 111, 98–121.
Tajfel, Henry (1981). *Human Groups and Social Categories: Studies in Social Psychology*. Cambridge: Cambridge University Press.
Wildschut, Tim, Brad Pinter, Jack L. Vevea, Chester A. Insko, and John Schopler (2003). "Beyond the Group Mind: A Quantitative Review of the Interindividual-Intergroup Discontinuity Effect." *Psychological Bulletin*, 129, 698–722.

Supplementary Material

Supplementary data are available at *JEEA* online.

Journal of the European Economic Association 2024 0(0):1–33
https://doi.org/10.1093/jeea/jvad072