Nastiness in Groups

By MICHAL BAUER, JANA CAHLÍKOVÁ, DAGMARA CELIK KATRENIAK, JULIE CHYTILOVÁ, LUBOMÍR CINGL, AND TOMÁŠ ŽELINSKÝ*

This version: November 31, 2020

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 on behalf of a group rather than when making choices individually on their own. We establish this behavioral regularity in four large-scale experiments among adolescents, university students and a nationally representative sample of adults (N = 7,426). We test several potential mechanisms, and the results suggest that the "destructiveness shift" in groups is driven by lower perception of individual responsibility, in line with self-signaling models.

(JEL: C92, C93, D01, D64, D74, D91)

* Bauer: CERGE-EI (a joint workplace of Charles University and the Economics Institute of the Czech Academy of Sciences), Politických vězňů 7, 111 21 Prague 1, Czech Republic, and Institute of Economic Studies, Faculty of Social Sciences, Charles University. Cahlíková: Max Planck Institute for Tax Law and Public Finance, Marstallplatz 1, 80539 Munich, Germany, and CERGE-EI. Celik Katreniak: City, University of London, Northampton Square, London EC1V 0HB, United Kingdom, and CERGE-EI. Chytilová: Institute of Economic Studies, Faculty of Social Sciences, Charles University, Opletalova 26, 110 00 Prague 1, Czech Republic, and CERGE-EI. Cingl: Faculty of Economics, University of Economics in Prague, nám. W. Churchilla 4, 130 67 Praha 3, Czech Republic. Želinský: Faculty of Economics, Technical University of Košice, Němcovej 32, 040 22 Košice, Slovakia. The data collections were funded by the Czech Science Foundation (P402/12/G130, Experiment 1 and 13-20217S, Experiment 2), Max Planck Institute (Experiment 3) and ERC-CZ/AV-B (ERC300851901, Experiment 4). Further support was provided by the Czech Science Foundation (Bauer and Chytilová, 17-13869S). We thank four anonymous referees, Gary Charness, Stefano DellaVigna, Dirk Engelmann, Armin Falk, Uri Gneezy, Shachar Kariv, Supreet Kaur, Martin Kocher, Filip Matějka, Jan Zápal, and participants at various seminars and workshops for helpful comments and conversations. An earlier version of this paper was circulated with the title "Anti-social behavior in groups". For correspondence: bauer@cerge-ei.cz; jana.cahlikova@tax.mpg.de; chytilova@fsv.cuni.cz.

Do people become nastier to other people when they act in groups rather than when they act individually on their own? Since writings about the limits of democracy by Plato (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] In his seminal work on individual behavior in crowds, Le Bon (1895, p. 35), for example, 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.[2] Some of this harmful behavior is hard to explain by pure calculated self-regard, and suggests that some people harbor a desire to be nasty and to harm others, even at personal cost to themselves. This paper provides evidence of the behavioral importance of the effect of making decisions in a group on greater prevalence of nastiness, and explores its sources.

Being nasty is a substantial step from being self-regarding. It refers to a desire to reduce another's material payoff for the mere purpose of harming, *without creating personal material gain and without fairness justification* (e.g., reducing inequality or a reciprocal response to hostility).[3] It is widely recognized that the prevalence of prosocial behaviors (reflecting concerns for fairness and the well-being of other people), at the expense of behavior guided by pure material self-interest, is important in determining a range of desirable societal outcomes – provision of

[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 being 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] Obstructionism refers to actions motivated by malicious intent that aim to interfere with an organization's ability to meet its objectives. Neuman and Baron (1998) describe the prevalence and various forms of obstructionism, including destruction of resources. For a summary of the literature on organizational workplace aggression, see Hershcovis et al. (2007).

[3] The literature uses different terms when describing the dark side of social preferences, meaning those preferences that place a negative weight on the well-being of other individuals, including: nastiness, spitefulness, anti-social behavior, aggressive competitiveness, and destructiveness. In this paper, we refer to this preference as nastiness or destructiveness and use these two terms interchangeably.

public goods, participation in public life, and cooperation (e.g., Fehr and Gachter 2002; Gintis et al. 2005). Experiencing pleasure from being nasty represents another fundamental -- but thus far less studied -- departure from own money-maximizing behavior (Abbink and Sadrieh 2009; Herrmann, Thoni, and Gachter 2008) 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, the scope for welfare losses and mutually destructive conflicts is magnified when decision-makers derive utility from harming others. Further, nasty behavior typically receives stronger condemnation than behavior guided by pure self-regard, as indicated by the term "sadism" used by social psychologists to describe behavior motivated by a negative weight placed on the well-being of others (Murphy and Ackermann 2014).

We conducted four large-scale experiments, all studying the influence of making decisions in a group context rather than individually, on the prevalence of nasty behavior. To measure such behavior, participants made decisions whether to financially harm an anonymous experimental counterpart at a small personal cost. We experimentally manipulate whether subjects make choices individually or on behalf of a group, but with no communication among group members. We refer to the difference in behavior as the effect of the group context. The initial set of lab-in-field experiments focused on adolescents in Slovakia (Experiment 1) and in Uganda (Experiment 2), and documents that making decisions in the group context systematically increases the prevalence of nasty behavior in both settings. Then, we probe further among a sample of university students in Slovakia (Experiment 3), in order to shed more light on the underlying mechanism of the observed effects. The results favor the interpretation that the behavioral change caused by the group context is driven mainly by reduced perception of individual responsibility. We also consider other mechanisms and show that motivation to please nasty group members, signaling toughness, action bias, and the role of in-group vs. out-group biases are unlikely to explain the "destructiveness shift" caused by the group context. Finally, in order to gauge the robustness of the main finding, we implement an online experiment among a nationally representative sample of adults in Slovakia (Experiment 4). We show that the effect holds across a broad spectrum of economic and demographic groups.

By providing the first causal evidence establishing how a group context influences the prevalence of nastiness, this paper contributes to existing experiments on decision-making in groups. Earlier work made important progress in studying situations 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. It has documented that groups or individuals in salient groups often behave less pro-socially than individuals do – they are less willing to sacrifice their own resources to increase social welfare or to achieve fair allocation of payoffs.[4] A prevailing interpretation of this stylized pattern is that decision-making in groups fosters rational self-regarding behavior, in line with textbook assumptions, meaning that groups are more prone to maximize their own payoff and to disregard the welfare of others (see excellent surveys by Charness and Sutter (2012) and Kugler, Kausel, and Kocher (2012)).[5] This "selfishness shift" is typically attributed to communication among group members, helping to recognize a dominant strategy. We show that the overall effect of decision-making in groups on lower prevalence of pro-social behavior is partly due to the effect of a group context making people more prone to engage in nasty actions, a "behavioral" channel identified in this paper. The analysis of additional measures of joint group decisions from Experiments 1-3 reveals that group deliberation and preference aggregation also matter, and foster a "selfishness shift", a mechanism featured in earlier research.

In the set of experiments described in this paper, subjects made decisions in the Joy of Destruction mini-game (Abbink and Sadrieh 2009; Abbink and Herrmann 2011). This task is designed to uncover the dark side of human social preferences. Subjects made a decision whether to pay a small amount of money in order to lower the reward of an anonymous counterpart.

[4] Most of this literature is based on comparisons of choices made by individuals and unitary teams/groups (typically composed of three members), in which group members make a joint team decision. Lower pro-social behavior among groups as compared to individuals has been documented in a range of economic experiments - the Dictator game, Ultimatum game, and the Trust game – see, for example, Kugler et al. (2007), Luhan, Kocher, and Sutter (2009) and Charness and Sutter (2012) for more references. Psychologists have extensively studied behavior in the Prisoners' Dilemma game; nearly all of their studies show that groups defect more often than individuals (see Wildschut et al. 2003 for a meta-study).

[5] This interpretation has powerful implications for economic theory. It suggests that group decisions can be modeled as more rational and less "behavioral" than individual decisions and that "*game theory based on standard assumptions may be, after all, a much better descriptive theory than currently believed*." (Kugler, Kausel, and Kocher 2012). Charness and Sutter (2012) summarize the emerging consensus as follows: "*The bottom line emerging from economic research on group decision-making is that groups are more likely to make choices that follow standard game-theoretic predictions, while individuals are more likely to be influenced by biases, cognitive limitations, and social consideration*."

Destroying the resources of a counterpart creates greater inequality, is not a response to unkind behavior, and is costly for the decision maker. Therefore, this task helps us to tease out whether people acting in groups destroy more due to greater self-interest or pleasure to be nasty, because the dominant strategy for purely self-regarding (or fairness-minded) individuals is not to engage in destructive behavior, whereas decision-makers that put a negative weight on the utility of the other person can engage in such harmful behavior. For ease of comparison of our results with earlier work, we also implemented the standard Prisoner's Dilemma game, a task that cannot separate desire to harm from self-interest, since both predict failure to cooperate.

The two initial experiments focus on the behavior of adolescents from disadvantaged regions. This subject pool is a natural starting point for this enquiry, as crime, urban riots and other forms of destructive behavior are more common among youth with lower socioeconomic status, as compared to other demographic groups (Lochner and Moretti 2004; Deming 2011). The experiments were conducted in schools in Eastern Slovakia (Experiment 1, N=630) and in rural Uganda (Experiment 2, N=1679). To elicit behavior in a group context, we asked subjects to state preferences about how they wanted their group of three individuals to decide when making a joint decision in the Joy of Destruction game (JDG) and in the Prisoner's Dilemma game (PDG), prior to discussing a joint decision with other group members. In addition, in Experiment 1, we elicited incentivized decisions, in which individuals made decisions on behalf of their group. We find that making a decision in a group context, rather than as an individual, increases the prevalence of destructive behavior in JDG. Similar effects of the group context hold in both settings, Slovakia and Uganda. In addition to observing the effects in JDG, we find that decision-making in the group context increases defection in PDG. Thus, the group context makes subjects more likely to harm others, both when such action brings them financial gains as well as when they need to sacrifice their own resources in order to do harm. We refer to this behavioral change in groups as a "destructiveness shift."

The destructiveness shift observed in Experiments 1 and 2 is consistent with several plausible mechanisms. First, self-signaling models highlight people's concern about building a positive self-image of being a moral person as an important regulator of inter-personal behavior (Benabou and Tirole 2011; Benabou, Falk, and Tirole 2018). Decision-making on behalf of groups involves reduced salience of self and creates perception of diffused responsibility, since more individuals are involved in the decision-making and can be attributed the responsibility for the

group decision (Darley and Latane 1968; Dana, Weber, and Kuang 2006). Both of these aspects are predicted to reduce self-image concerns, allowing people with a latent intrinsic preference to be nasty to more freely act on such desires.[6] Second, work on group identity suggests that salient group boundaries may give rise to "in-group/out-group" biases and lead to aggressively competitive behavior, with groups aiming to out-compete other groups in relative terms (Tajfel 1981; Durlauf 1999). Finally, people may care about their social image and what other group members think of them (Benabou and Tirole 2006; Bursztyn and Jensen 2017). When choices on behalf of a group are observable by other group members (or are expected to be discussed with peers), some individuals may choose to act destructively, in order to signal toughness to others or to please anti-social group members.[7]

Therefore, as a next step, we conducted Experiment 3 among 795 university students from Eastern Slovakia, which aims to isolate the role of reduced perception of individual responsibility in groups from other mechanisms. We implemented two experimental conditions in which subjects made decisions on behalf of a group. They were matched with two other individuals, knowing that one group member's decision would be randomly selected and treated as the decision of the whole group. This feature is predicted to create a perception of diffused responsibility and "plausible deniability", making it easier to express intrinsic preferences that violate social norms. In a *GroupContext_Hidden* condition, other mechanisms should not play a role. To close the effect of social image considerations, subjects did not know who the other group members were, did not interact with them in any way, and did not observe each other's decisions. In-group/out-group biases were unlikely to play a role either, because group members were completely anonymous to each other and did not share any specific group attribute. In contrast, in a *GroupContext_Observed* condition, subjects knew that their choices would be observed by the other group members, allowing social image considerations to play a role, and they made choices when sitting next to each other, making the group boundary salient.

[6] A related idea from psychology is termed "deindividuation" and describes a reduction of inner restraints when individuals are "submerged in a group" (Diener 1980). When individuals become part of a group, they tend to be less self-aware and lose their individualised sense of identity, thus loosening their normal inhibitions that limit aggressive and deviant behavior.

[7] A priori, the predicted effects of social image concerns on the prevalence of nasty behavior are unclear, since such effects depend on whether people aim to signal their kindness or nastiness to others. Note that, empirically, the existing evidence suggests that most people care about a positive social image and behave more kindly when they are being observed by others (see, for example, Andreoni and Bernheim (2009) or DellaVigna, List, and Malmendier (2012).

We replicate the main finding identified among adolescents and find the following patterns. First, making decisions on behalf of a group in *GroupContext_Hidden* increases the prevalence of destructive behavior in comparison to making individual decisions. Second, making decisions in *GroupContext_Observed* also increases the prevalence of destructiveness, compared to making individual decisions, but the prevalence is similar as (or slightly lower than) in *GroupContext_Hidden*. Together, these patterns provide support for the interpretation that contextual factors that reduce perception of individual responsibility are the main driver of the increased destructiveness when individuals make choices as a part of a group. They also suggest that observability and a salient group boundary are unlikely to play major roles in causing the "destructiveness shift" in groups. Further, data on beliefs about the preferences of other group members do not support the idea that the increase in nastiness is driven by effort to please anti-social individuals within one's own group.

In Experiment 4, we collaborated with a survey agency (IPSOS) and implemented a large online experiment on a nationally representative sample of adults in Slovakia (N=4,243). Participants made decisions individually or in *GroupContext_Hidden*. We again replicate the main pattern – making decisions on behalf of the group systematically increases the prevalence of destructive behavior in JDG and also increases the prevalence of defection in PDG. Further, we show that the effect holds for various sub-groups, in terms of gender, age, education, income level, size of municipality, and political orientation of the respondents. Taken together, Experiments 1-4 provide evidence of a systematic behavioral regularity: making decisions on behalf of a group, rather than as an individual, magnifies the prevalence of nasty behavior, and this result holds across a broad range of demographic and economic groups.

Finally, we show that qualitatively different mechanisms give rise to the "destructiveness shift" and the "selfishness shift" in groups. In addition to making decisions on behalf of a group, in Experiments 1-3 subjects also made decisions as unitary groups composed of three members. The groups were asked to deliberate and to make a joint decision. Comparing joint group decisions and individual decisions on behalf of a group allows us to estimate the effect of group decision-making (group deliberation and preference-aggregation), and to contrast it with the effect of a group context. We find that group decision-making also influences behavior. In all three experiments, group decision-making increases the prevalence of defection in PDG, while the prevalence of destructive behavior in JDG is either reduced (among university students) or remains

at a similar level as when individuals make decisions on behalf of a group (among adolescents). Thus, the results are broadly supportive of the interpretation that deliberation and preference aggregation in unitary groups increases the prevalence of own (group) money-maximizing behavior.

This paper is related to several streams of literature. First, it adds to emerging literature that has documented that nasty behavior is not an anomaly, committed by a narrow set of individuals from the fringes of society (Falk, Fehr, and Fischbacher 2005; Herrmann, Thoni, and Gächter 2008; Abbink and Herrmann 2011; Prediger, Vollan, and Herrman 2014). In all experiments we implemented, the lowest prevalence of choosing the destructive option in JDG is 12%, among university students. Among the representative sample of adults, we have not found any demographic sub-group for which the prevalence of destructive behavior has been less than 15%.

Second, the paper relates to literature on "moral wiggle room" and perception of individual responsibility. Laboratory studies have documented that contextual factors, that provide scope for excuses by obscuring the role of the decision-maker in determining an outcome (Dana, Weber, and Kuang 2006; Dana, Loewenstein, and Weber 2011), reduce the likelihood of making a pivotal decision (Falk, Neuber, and Szech 2020) or reduce salience of self (Falk 2017) have substantial effects on behavior, and 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 the behavioral model of self-signaling (Benabou and Tirole 2011; Benabou, Falk, and Tirole 2018), in which environmental factors that reduce perception of individual responsibility allow individuals to maintain a positive self-image even when acting immorally. Our findings suggest that acting on behalf of a group, rather than on behalf of one's self, is another important and widespread decision environment in which perceptions of reduced individual responsibility systematically affect behavior. In addition, while much of the literature features the role of self-image costs in regulating the prevalence of purely selfish behavior, our results suggest that they may keep purely destructive inclinations at bay as well.

Finally, this paper helps to explain why decisions in groups and joint group decisions are often different from individual decisions in the inter-personal domain (e.g., Bornstein and Yaniv 1998; Wildschut et al. 2003; Kugler et al. 2007). While the novel finding of this paper is that a

group context may foster nastiness, we also document a shift towards self-regarding behavior due to group deliberation and preference-aggregation, especially among university students, and thus provide further support for the mechanism highlighted in earlier body of work (Charness and Sutter 2012; Kugler, Kausel, and Kocher 2012).[8] To do so, we use an integrated experimental design that allows us to separate the impacts of group context and group deliberation. Our design is inspired by Sutter (2009) and Feri, Irlenbusch, and Sutter (2010), who use a related experimental design to study decision making under uncertainty and ability to coordinate, and it helps to bridge the literatures on the effects of group context, pioneered by Charness, Rigotti, and Rustichini (2007), and a larger literature on joint decision-making in groups (e.g., Ambrus, Greiner, and Pathak 2015; Balafoutas et al. 2014).

The remainder of this paper proceeds as follows. In Section I, we present the design and results from Experiments 1 and 2, which study behavior in groups among adolescents. In Section II, we present the design and results of Experiment 3, which studies the role of different mechanisms behind why a group context increases nastiness among university students. In Section III, we present the design and results of Experiment 4, which explores the robustness of the main findings among a diverse sample of adults. Section IV concludes.

I. EXPERIMENTS 1 AND 2: EVIDENCE AMONG ADOLESCENTS

The aim of the two initial studies was to identify whether individuals behave more destructively towards others when they make choices in a group context than when they decide individually in isolation. In addition, we were interested in whether the observed effects persist when group members make a decision as a group, after joint deliberation and preference aggregation. Thus, Experiments 1 and 2 were designed to separately identify the effect of a group context and the effect of group decision-making. To establish the initial patterns, we focused on a population of

[8] The economic literature on group decision-making also relates to the literature on "discontinuity effects" in social psychology. Discontinuity effects refer to an empirical pattern identified in the Prisoners' Dilemma game, namely that when groups interact with groups they are less likely to cooperate than when individuals interact with other individuals (Schopler and Insko 1992; Wildschut et al. 2003). Psychologists have interpreted the evidence as an indication of greater self-regard and/or greater competitiveness and aggressive behavior in groups (Wildschut et al. 2003). Nevertheless, since the evidence rests on choices in PDG and similar tasks, the role of own-money maximizing motivations and aggressive behavior (harming others at a cost to one's self) have remained an open question.

adolescents for whom the prevalence of destructive inclinations might be expected to be relatively high, given that crime, vandalism and other forms of anti-social behavior are more common among youth (Lochner and Moretti 2004).[9]

A. Sample selection

The initial experiments were implemented in 2013 among adolescents in two countries – the Slovak Republic, an OECD and European Union member country, and Uganda, a low-income country in sub-Saharan Africa. Experiment 1 was implemented among a sample of 709 participants from the last two grades of 13 primary schools located in villages and small towns scattered across Eastern Slovakia (see the map in Figure A.1). The participants were ages 13–15. Experiment 2 was implemented among a sample of 1,679 participants from the last grade of 24 primary schools and the two final grades of 10 lower-secondary schools in rural areas of southern Uganda (Figure A.1). Because many students repeat grades, there is substantial age variation within the Ugandan sample (90% of subjects were ages 12-18).

Our decision to conduct the experiments in Slovakia and Uganda was guided by two considerations. First, we aimed to test the robustness of the main patterns across two very different settings, in light of growing concerns about replicability of experiments (Maniadis, Tufano, and List 2015). Second, our team had an established data collection infrastructure in these two specific settings.

A comparison of sample characteristics from Experiments 1 and 2 (Tables A.1 and A.2) documents the diversity of living conditions across the two settings. Slovak subjects live in standard developed-country housing, and 98% of their families own a TV and 89% own a car. They have on average two siblings, and almost all of their parents have a secondary or tertiary degree. In contrast, the Ugandan subjects typically live in housing with no electricity (62%), their families are much larger, with seven siblings on average, and 34% of mothers and 25% of fathers have primary school education or less.

[9] This link between age and destructive behavior is also observed in our experiments. We find that adolescents (Experiments 1 and 2) tend to be more destructive than adults (Experiments 3 and 4), and that younger individuals tend to be more destructive than older individuals (Experiment 4, Figure 6).

B. Experimental tasks

The main task we implemented is the Joy of Destruction mini-game (JDG) (Abbink and Herrmann 2011; Abbink and Sadrieh 2009), a money-burning game designed to identify destructive preferences. Two players received the same endowment (EUR 2 in Experiment 1 and USh 1,000 in Experiment 2) and simultaneously decided whether to sacrifice 10% of their endowment in order to decrease their counterpart's payoff by 50% (destructive choice) or whether to keep the payoffs unchanged (non-destructive choice). Thus, the participants made an unconditional decision, i.e. chose an action without knowing what the other player 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 participant's and 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, reciprocity or inequality aversion. Because the decision is one-shot and the destructive choice is costly in the sense that the participant has to sacrifice a part of own payoff, the dominant strategy of a purely self-regarding player is not to engage in destructive behavior. Sequential fairness motives, such as reciprocity, also cannot justify destruction, since choices were made simultaneously without knowing what the counterpart did, and thus were not made in response to an unkind counterpart's behavior.[10] Inequality aversion cannot explain destructive behavior because the destructive action does not diminish inequality, but increases it.

In order to measure the willingness to cooperate, and for comparison with earlier work, we also administered the Prisoner's Dilemma game (PDG). Two players received the same endowment (EUR 1.6 in Experiment 1 and USh 800 in Experiment 2) and simultaneously decided whether to take away 50% of their counterpart's payoff in order to increase their own payoff by 25% (defection/non-cooperative choice) or whether to keep the payoffs unchanged (cooperative choice). Defection is a dominant strategy for both purely self-regarding players and for players

[10] In principle, destructive choice in unconditional decision in this task could be motivated by negative reciprocity combined with a belief that the experimental counterpart is destructive. We show that the main effect of the group context is unlikely to be driven by these beliefs, by documenting in Experiments 3 and 4 that the effect also holds for the conditional decisions.

with destructive preferences, but the socially optimal outcome is reached when both players cooperate.

The payoff matrices for both JDG and PDG as administered in Slovakia and in Uganda are presented in Figure 1 (Panels A and B). We used neutral framing. For example, in Experiment 1, the question in JDG was "Do you want to pay 20 cents and reduce the other person's income by 1 euro?" In PDG, we asked "Do you want to take 80 cents from the other person to get 40 cents for yourself?" The order of JDG and PDG was randomized.[11] The counterparts in the games were always anonymous. In Experiment 1, the counterparts came from an unspecified school in the same region, and thus were completely unknown to the decision makers. In Uganda, the counterparts were decision-makers' classmates whom they had known for several years.

C. Experimental manipulations

The experimental manipulations were designed to identify two types of group effects on destructiveness: (i) the effect of group context on individual decisions and (ii) the effect of group decision-making, i.e. of group deliberation and preference aggregation. Consequently, Experiments 1 and 2 contain three conditions in which participants made choices (a) individually (*Individual* condition), (b) individually in a group context (*Group Context* or *GC* condition), (c) jointly as a group (*Group* condition).

In the *Individual* condition, subjects made choices in isolation and did not interact with other participants in any way. The counterpart also made choices individually.

In the *Group* condition, subjects were randomly allocated into groups of three. They were matched with real-life peers from their class, with whom they regularly interacted. When making choices, they sat next to each other around one desk (Slovakia) or on one mat (Uganda) and were informed that they had up to four minutes of free discussion to reach a joint decision. It was specified that the payoffs described in the instructions were per member and that each member of the group would receive the same payoff.

[11] In Study 2, we further implemented the Competitiveness task (Niederle and Vesterlund 2007) and find that a group context increases willingness to compete. We think this is an interesting result on its own, but, conceptually, it is not clear why destructive preferences should motivate people to compete in this task. In line with this, we also find that the correlation between behavior in the Competitiveness task and in JDG is relatively low (0.059). Thus, we do not feature this result in the main paper and describe the findings in the Online Appendix.

In order to separate the effect of the group context on individual decisions and the effect of group decision-making, subjects in the *Group* condition were asked to state their individual preference regarding the group decision (*GroupContext_1* or *GC_1* condition), prior to the group deliberation stage. The participants made these choices in private, but they could expect communication about their decisions with other group members after completing the task because they knew they would discuss the choice and make a joint decision in the *Group* condition.

The *Individual* and *GC_1* conditions were implemented using a between-subject design. By comparing the behavior in *GC_1* and in *Individual*, we estimate the effect of the group context on individual decisions. The *GC_1* and *Group* conditions were implemented using a within-subject design. Comparison of the behavior in *Group* and *GC_1* allows us to estimate the effect of group decision-making.

In Experiment 1, we further implemented a second version of the group context condition (*GroupContext_2* or *GC_2)*, in order to study the robustness of the observed effects when using an alternative measure of decision-making in the group context. Subjects were again matched with two classmates. Instead of expressing their preference regarding the group decision, they made decisions individually on behalf of their group knowing that one group member's decision would be randomly selected and treated as the decision of the whole group. The choices were observed by the other two group members.[12]

In addition, in Experiment 1, we included sub-conditions to explore whether the observed effects of the group context arise in a broad range of interactions, including when destructive behavior impacts isolated individuals or individuals from participant's own communities, or whether the group setting makes people more destructive only when interactions involve other teams or members from socially distant groups. Specifically, we manipulated whether the counterpart was of the same (majority) or different (Roma) ethnicity[13], and in the *GC_1* and *Group*

[12] In a separate paper, Bauer et al. (2018), we study how the choices made by subjects in *GC_2* influence the choices of the other two matched peers who made their own choices afterwards, in order to estimate the level of contagion of hostility. Here, to avoid confounds due to such peer effects, we analyze only the choices of the subject who made the decisions first (i.e., without observing the choices of their peers' prior own decision). The results are similar when we include the choices of subjects who made decisions second and third (Table A.16).

[13] These two conditions were implemented using a within-subject design, in a random order. Because we observe very similar effects (see Table A.6 and the discussion in Section I.E.2), in the main estimates we pool observations from these two conditions and cluster standard errors at the individual level. Ethnicity was signaled using a list of 20 real names of potential counterparts; the list contained either typical Slovak names or typical Roma names.

conditions, we manipulated whether the counterpart was an individual or another group of individuals.

We do not find systematic differences in observable characteristics across conditions, indicating that the randomization was successful (Tables A.1 and A.2).

D. Procedures

The experiments were implemented in schools. The tasks were explained using visual aids to illustrate the options and payoffs. Before making choices in JDG, the participants were asked four control questions about the payoff consequences of their actions and their counterpart's actions. The same procedure was followed for PDG. The decisions were made anonymously, by submitting answers under an experimental ID.

The choices were incentivized, except in the *GC_1* condition, in which we asked hypothetical questions. The subjects were paid for a subset of randomly selected decisions. In Experiment 1, subjects received rewards in the form of credit to order from a variety of items available in an experimental store, because headmasters of the schools requested that we not use monetary rewards. In Experiment 2, subjects received monetary rewards. To avoid communication about experimental tasks prior to participating, all subjects from each class participated in the experiment at the same time and all sessions within each school were implemented in a single day. Online Appendix B contains the full experimental protocol.[14]

E. Results

E.1 Overall levels of destructive behavior

We find that a relatively large percentage of subjects in *Individual* chose to destroy in JDG. The prevalence of destructive choices in *Individual* is 32% in Slovakia and 53% in Uganda. Other studies that have implemented money-burning games among more general samples of population

[14] Since at the time when Experiments 1 and 2 were implemented, the institutions the authors were affiliated with did not have official IRB, these two studies were approved by the Director of the Institute of Economic Studies at the Faculty of Social Sciences, Charles University, and by the headmasters of participating schools. Study 1 was further approved by the Dean of the Technical University of Košice and Study 2 was officially supported by the Buikwe District Local Government in Uganda. Participation in the experiments was voluntary and the subjects could leave at any time.

than selected samples of university students also indicate that destructive behavior is quite prevalent. For example, among adults, destruction rates were 20% in Kenya (Bauer, Chytilová, and Miguel 2020), 19% in Slovakia (Experiment 4 in this paper), 30% in India (Fehr, Hoff, and Kshetramade 2008) and 23% in Namibia (Prediger, Vollan, and Herrman 2014), and higher, depending on the experimental conditions (see Table A.3 for a more detailed comparison). Thus, the destruction rates among adolescents observed in our Experiments 1 and 2 are somewhat higher than in adult samples.

The level of understanding of the games was high -- 88% of individuals and groups in Slovakia and 70% in Uganda answered all comprehension questions correctly. The results are robust to excluding observations with imperfect understanding (Table A.4 and Table A.5).

E.2 Effect of group context

In order to identify the effects of group context on individual behavior, we compare individual choices made in isolation (*Individual*) with the preferences of individual group members on how they wanted their group to decide (*GroupContext_1*).

In the Joy of Destruction game, we find that deciding in a group context magnifies the prevalence of destruction (Figures 2 and 3; Panel A of Tables 1 and 2). In Slovakia, the prevalence of destructive choices increases from 32% in *Individual* to 44% in *GC_1* (p-value = 0.004) and to 45% in *GC*_2 (p-value = 0.026). In Uganda, subjects in *Individual* chose to destroy in 53% of cases, compared to 60% in the *GC_1* condition (p-value = 0.012).[15] The results in both countries are robust to controlling for the order of the games, experimenter, school and school-grade fixed effects, and a range of observable characteristics (Tables A.4 and A.5). Thus, in both countries, when adolescents made decisions in a group context rather than individually, they were less likely to play the Nash equilibrium consistent with self-regarding preferences, and more likely to choose a purely destructive strategy.

Next, we study behavior in the Prisoner's Dilemma game (Panel B of Tables 1 and 2). In Slovakia, the prevalence of defection increases from 67% in *Individual* to 75% in *GC_1* (p-value

[15] In the analysis, we treat the choices of three members matched into a group in *GC_1* as independent. However, the subjects knew with whom they were matched and that they would later communicate about the *Group* decision. Therefore, as a robustness check, we allow for clustering at the group level. The results are unchanged (Column 9 of Tables A.4 and A.5).

= 0.035) and to 72% in *GC_2* (p = 0.436). In Uganda, subjects in *Individual* chose to defect in 57% of cases, compared to 68% in *GC_1* (p-value < 0.001).

Several additional results illuminate the robustness of the pattern observed. First, social distance between the decision-maker and the experimental counterpart has virtually no influence on whether subjects behave more destructively in a group context. The results are similar in Slovakia, where the counterparts were completely unknown to the decision makers, and in Uganda, where the counterparts were the decision-makers' classmates. A group context also increases the prevalence of destructive behavior independently of the counterpart's ethnicity (Experiment 1, Panel A of Table A.6). Finally, the shift in destructiveness caused by the group context occurs both when the counterpart is a group and when the counterpart is an individual (Experiment 1, Panel B of Table A.6).

E.3 Effect of group decision-making

In the previous section, we show that a group context magnifies destructiveness by comparing behavior in *GC_1* and *Individual*. In this section, we explore whether group decision-making (involving group deliberation and preference aggregation) attenuates or further amplifies this effect, by comparing the behavior in *Group* and *GC_1*.

In both countries, we find that group decision-making has virtually no influence on the prevalence of destructive behavior in JDG (Figures 2 and 3, Panel A of Tables 1 and 2). In Slovakia, the prevalence of destructive choices is 44% in *GC_1* and 42% in *Group* (p-value = 0.482). In Uganda, the prevalence of destruction is 60% in *GC_1* and 59% in *Group* (p-value = 0.910). Thus, the greater tendency to act destructively in a group context, as compared to when acting individually, persists even when members make a joint decision as a whole group. Consequently, the proportion of destructive choices is significantly larger in *Group* than in *Individual* (p = 0.055 in Slovakia and p=0.052 in Uganda, Tables 1 and 2). At the same time, the finding that group deliberation does not further elevate destructiveness indicates that destructive members of the group are not disproportionately influential during the group deliberation, and thus that this mechanism does not contribute to greater nastiness in groups.[16]

[16] This conclusion is further corroborated in Tables A.17 and A.18. One way to gauge how influential different behavioral types are is to analyze how composition of a priori preferences of individual group members predicts the

Unlike in JDG, in PDG group decision-making influences the prevalence of non-cooperative behavior (Panel B of Tables 1 and 2). In both countries, we find that the individual preferences for the group decision (*GC_1*) are located almost exactly in between the choices made by individuals in isolation (*Individual*) and the group choices (*Group*). Thus, group decision-making causes additional increase in the prevalence of defection, beyond the effect of the group context on individuals.

In the analysis above, we focus on unconditional choices, which we elicited in all experimental manipulations. In the *Individual* and *Group* conditions, we further elicited two conditional decisions - for the situation when the counterpart decided to keep the payoffs unchanged, and for the situation when the counterpart decided to lower the decision maker's payoff. Since in the conditional decisions the participants knew the action of their counterpart, beliefs about the counterpart's decision should be irrelevant, and thus decisions should be informative about the participant's preferences. The differences in behavior between *Group* and *Individual* hold for six of eight conditional decisions (Tables 1 and 2). When the counterpart is non-destructive (in JDG) or cooperative (in PDG), groups are more destructive/less cooperative in both countries. In Uganda (but not in Slovakia), the effect also holds when the counterpart is destructive (resp. non-cooperative). Further, we find no systematic differences in beliefs about a counterpart's behavior (elicited as a binary variable using hypothetical questions) across *Group* and *Individual* (Tables 1 and 2). Together, the results suggest that behavioral differences are not driven by changes in these beliefs.

ultimate group decision, since subjects in the *Group* condition first stated their preferences about the group decision (in the *GC_1* condition) and then participated in making joint decisions. Note that we cannot perform a similar analysis for subjects allocated in the *GC_2* condition, since they did not make joint group decisions. We find that in groups where they represent a minority opinion, destructive types are not more influential than non-destructive types. Specifically, groups with one destructive member (and two non-destructive members) are *not* more likely to switch from behavior predicted by median opinion and end up choosing the destructive option, as compared to the likelihood that groups that have one non-destructive member (and two destructive) end up choosing the non-destructive option.

E.4 Summary and open questions

To summarize, in both settings we study, we find that letting adolescents to make decisions in a group context, rather than individually in isolation, makes them more prone to sacrifice their own resources to financially harm someone else. Further, group decision-making has virtually no influence on the prevalence of destructive behavior. At the same time, we find some indication of an increase in self-regarding behavior at the expense of cooperative behavior caused by group decision-making, because in PDG, joint group decision-making causes additional increase in the prevalence of defection, beyond the effect of the group context.

Although our experiments with adolescents document the important role of a group context in elevating the prevalence of destructive behavior in this population, they also leave several important questions open. The first is whether the effects generalize to adults, arguably a more important population since it can engage in destructive behavior on a larger scale. In order to study this question, we implemented follow-up experiments in Slovakia with university students (Experiment 3) and a representative population of adults (Experiment 4).

Second, the design of Experiments 1-2 cannot separate several potential mechanisms through which a group context may increase the prevalence of destructive behavior. First, when individuals make decisions on behalf of a group rather than on behalf of themselves, they may perceive themselves as being less responsible for the decisions, reducing costs to self-image from deviating from social norms that prohibit destructive behavior. Therefore, intrinsic preferences may manifest more easily. Second, social image considerations may matter, too. Specifically, subjects may see value in signaling toughness to peers by acting destructively. In *GC_2*, choices made on behalf of a group were directly observed by other group members. In *GC_1*, choices were made in private, but subjects expected future communication with group members about the decisions. If they were lying averse, social image considerations could still be relevant and the subjects could have adjusted their stated preferences. Third, if subjects thought both other group members preferred the destructive strategy, they may have chosen the destructive option in order to please them. Finally, in *GC_1* and *GC_2* conditions, subjects sat in close proximity to other group members (around a table or on a mat), making the boundaries between groups salient. The literature on minimal group experiments suggests that sharing a group attribute or physical proximity may trigger a greater sense of shared group identity (Tajfel and Turner 1979), potentially

leading to more aggressively competitive behavior towards individuals outside of one's own group.

II. EXPERIMENT 3: TOWARDS UNDERSTANDING THE MECHANISMS OF THE EFFECTS OF A GROUP CONTEXT

The first follow-up study, Experiment 3, was implemented among a sample of university students. The main aim was to shed light on the potential mechanisms of group context effects described above: reduced perception of individual responsibility, signaling to others, pleasing peers, and in-group/out-group biases. To do so, we took several steps.

We added a new experimental condition, *GroupContext_Hidden* (*GC_Hidden*). This condition is designed to retain only those features of the group context that are predicted to affect perception of individual responsibility for the decision, while isolating the role 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 feature can create a perception of diffused responsibility since the subjects may exploit uncertainty about precisely who caused the nasty outcome. Also, the decision was framed as a decision on behalf of the group, rather than as an individual decision, which may lower self-attribution of responsibility. At the same time, in order 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. Further, in-group/out-group biases are also unlikely to play a role in this condition, since this decision environment did not involve a salient group boundary. Members of the same group were completely anonymous to each other, were not in physical proximity 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 the perception of diffused responsibility.

The second condition, in which subjects made choices on behalf of a group, is *GroupContext_Observed* (*GC_Observed*). It is similar to the *GC_2* condition implemented among

adolescents in Slovakia. Members of the same group made choices while sitting next to each other, rendering the group boundary salient. They knew that their choices would be observed by the other group members, allowing signaling and social image considerations to play a role. Thus, we interpret the difference in behavior between *GC_Observed* and *GC_Hidden* as capturing the combined effect of signaling and salient group identity.

Next, Experiment 3 improves measurement of the outcome variables in several ways. First, all the choices were incentivized. Second, in addition to unconditional decisions, we elicited two conditional decisions and beliefs about the behavior of the counterpart in all experimental manipulations (in Experiments 1 and 2 we elicited these measures only in *Individual* and in *Group*). This allows us to address the question of whether the group context affects destructive preferences, rather than beliefs about a counterpart's behavior, since, in principle, the destructive choice in the unconditional decision can be motivated by beliefs about the likelihood of destructive behavior by the counterpart combined with negative reciprocal preferences. In contrast, in conditional choices, beliefs about a counterpart's behavior should not play any role.

Finally, we elicited individual beliefs about the behavior of other group members, in order to gauge whether subjects expect them to prefer the destructive strategy. This helps us to address the question of whether the effects of making decisions on behalf of a group could originate in motivation to please nasty group members.

A. Sample

Experiment 3 was implemented among 795 university students from Eastern Slovakia. The participants were students from the Technical University of Košice and the University of Prešov. Since there is no experimental economics laboratory in the region, the experiments were organized during lectures or 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.

B. Experimental tasks

The main task we implemented was again the Joy of Destruction mini-game (JDG). Two players received EUR 10 each, and chose whether to pay EUR 0.5 to destroy EUR 5 of the counterpart or whether to keep the payoffs unchanged. Previous evidence suggests that the prevalence of destructive behavior tends to be lower among subjects with higher socio-economic status (Bauer, Chytilová, and Pertold-Gebicka 2014; Prediger, Vollan, and Herrman 2014).[17] In order to have enough variation in the measure of destructive behavior, destruction was relatively cheaper in Experiment 3 than in the first two experiments—subjects had to pay 5% of the endowment (instead of 10%) to destroy 50% of their counterpart's payoff. In the Prisoner's Dilemma game, both players received EUR 8 and decided whether to take EUR 4 from the counterpart to gain EUR 2 for themselves, or whether to keep the payoffs unchanged. Panel C of Figure 1 displays the payoff matrices for JDG and PDG.

In both JDG and PDG, we elicited unconditional decisions and two conditional decisions – one for the situation when the counterpart decided to keep the payoffs unchanged, and one for the situation when the counterpart decided to lower the decision maker's payoff. We also elicited beliefs about the behavior of the experimental counterpart, by asking subjects to guess which of the two strategies was selected by the counterpart. Thus, the measure of beliefs is a binary variable. Further, we also elicited beliefs regarding the unconditional choice of the other two group members (*GC_Hidden* and *GC_Observed*).[18]

We also implemented two additional tasks in this experiment. First, in order to assess whether differences in destructive behavior when subjects make decisions jointly as a group vs. individually could be driven by a greater urge to do something (as opposed to adhering to the status quo) when being in a group, we implemented an "action bias" task in *Individual* and *Group* conditions. We do not find any significant differences in choices in this task across the two conditions. We describe this task and the results in more detail in Section III, along with the findings from Experiment 4. The second additional task measured the willingness of participants

[17] This correlation is also confirmed in Experiment 4, which shows that more educated individuals are less likely to destroy in JDG (Panel D of Table A.14).

[18] In order to follow the same experimental procedure and incentive structure across conditions, in *Individual* we elicited beliefs about the behavior of two other participants in the same session.

to destroy a monetary donation to their university, an even more extreme manifestation of destructiveness than JDG. The patterns are qualitatively similar as in JDG (except in *GC_Observed*), but the prevalence of destructive behavior in this task was very low in all treatments (0.02 (5 individuals) in *Individual*, 0.03 (7 individuals) in *GC_Hidden*, 0.004 (1 individual) in *GC_Observed* and 0.01 (2 individuals) in *Group*). Further details are provided in the Online Appendix.

C. Experimental conditions

Subjects were randomly allocated into one of the following three conditions, using a between-subject design: *Individual*, *GC_Observed* and *GC_Hidden*. In *Individual*, subjects made choices individually. In *GC_Observed*, subjects decided on behalf of a group of three members. Each subject was randomly matched with two participants from the same session and all group members sat next to each other. The participants were informed that a decision made by one randomly selected group member would be applied for the whole group. The choices were directly observed by the other two group members. In *GC_Hidden*, subjects also decided on behalf of a group of three, but the three group members were anonymous to each other. Participants knew that the other two group members were participants from another day and that they could not interact with them in any way. Observable characteristics of participants vary little across the experimental conditions (Table A.7).

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 four minutes to reach a joint group decision. Thus, the *Group* condition came as a surprise, and the expectation of group decision-making should not have influenced the choices made in the other conditions. Subjects were matched (*Individual*) or re-matched (*GC_Observed* and *GC_Hidden*) into groups of three with two other participants from the same session.

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, i.e. in *Individual* the counterpart was another individual, in *GC_Hidden* and *GC_Observed* the

counterpart was a group of three participants who made decisions on behalf of their group, and in *Group* the counterpart was another group of three participants who made joint decisions.

D. Procedures

As in Experiments 1 and 2, the tasks were explained using visual aids and the participants were asked four control questions about the payoff consequences of their actions and their counterpart's actions before making choices in JDG and in PDG. The level of understanding was high. In both games, 95% of subjects answered all control questions correctly, and the results are robust to excluding observations with imperfect understanding (Table A.8).

The decisions were made anonymously, by submitting answers 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 (EUR 2 for a correct guess). The subjects received rewards in cash. After the session, the participants received EUR 5, knowing they would receive the rest of their payoff after all decisions were matched, by coming to one of "payout" sessions at the campus and showing their experimental ID card. Specific instructions for the three experimental conditions were given separately in three rooms. Online Appendix B contains the full experimental protocol.[19]

E. Results

E.1 Effects of a group context

In the analysis, we first use an index that captures all decisions about whether to choose the destructive option in JDG. The index is constructed as an average of three indicator variables for whether a decision-maker acted destructively in the unconditional choice and in two conditional

[19] Experiment 3 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. Participation in the experiment was voluntary and the subjects could leave at any time.

choices, and thus takes a value between zero and one. Then, we proceed with discussion of the effects of the group context on unconditional and conditional choices separately.

We find that, as with the earlier experiments, making decisions on behalf of a group also systematically increases the prevalence of destructive behavior in JDG among university students (Figure 4 and Panel A of Table 3), replicating the main pattern observed among adolescents. The prevalence of destructive behavior increases from 20% in *Individual* to 27% in *GC_Hidden* (p-value = 0.001). Interestingly, this effect is similar but somewhat smaller in magnitude in *GC_Observed*: the prevalence of destruction is 24%, which is larger than in *Individual* (p-value = 0.051) and (marginally) not statistically significantly different from *GC_Hidden* (p-value = 0.180). The results are robust to controlling for the order of the games, experimenter fixed effects, and a range of observable characteristics (Table A.8).

The effects are similar when we focus on unconditional and conditional decisions (Table 3). For the unconditional decision, the prevalence of destructive behavior increases from 12% in *Individual* to 18% in *GC_Hidden* (p-value = 0.090), while in *GC_Observed* it is 12%. Conditional on an experimental counterpart choosing the non-destructive strategy, a situation in which it is particularly nasty to opt to destroy, 4% of subjects in *Individual* chose to destroy, and the destruction rate increases to 13% (p-value < 0.001) and 7% (p-value = 0.143) in *GC_Hidden* and *GC_Observed*, respectively. The effects are qualitatively similar when the counterpart is destructive: 43% of subjects in *Individual* chose to destroy, and this increases to 51% (p-value = 0.083) and 54% (p-value = 0.019) in *GC_Hidden* and *GC_Observed*, respectively.

In PDG, making decisions on behalf of a group also matters and leads to higher defection rates, although the effects are smaller in magnitude than in JDG, and only marginally statistically significant (Panel B of Table 3). When looking at the index constructed from all three choices, the defection rate is 59% in *Individual* and 64% in *GC_Hidden* (p-value = 0.068) and in *GC_Observed* (p-value = 0.082). This effect is mainly driven by the conditional decision in which subjects respond to a situation when an experimental counterpart acted cooperatively.

E.2. Which type of behavior becomes more prevalent in the group context?

In this sub-section we consider whether a group context intensifies pure destructiveness, or whether the effects can be explained by greater selfishness, changes in beliefs or a stronger preference for retaliation.

As described earlier, destructive behavior in JDG is one-shot, anonymous, and costly for the decision-maker. Thus, by design, the higher prevalence of destructive behavior in the group context cannot be explained by increased selfishness.

The effects are unlikely to be driven by changes in beliefs about the actions of the experimental counterpart either, specifically by greater concern that the counterpart will behave destructively in a group context. While we observe that the proportion of subjects who believe that their counterpart behaved destructively is higher in *GC_Hidden* (24%) than in *Individual* (15%), this is not the case in *GC_Observed*, where the proportion is the same as in *Individual* (15%). Thus, changes in beliefs cannot explain why destructiveness increases in *GC_Observed*, as compared to *Individual*. More importantly, the group context increases the prevalence of destructive behavior not only in the unconditional decisions, but also in both conditional decisions, in which beliefs about a counterpart's behavior should not play any role (Table 3).

There are two qualitatively different types of destructive preferences. The first is a retaliatory response to unkind behavior of others; in some circumstances such negative reciprocity may help to sustain cooperation in the long-run (Gächter, Renner, and Sefton 2008). The second is pure nastiness, which urges people to cause harm to completely passive or cooperative individuals; such behavior undermines cooperation even beyond the level predicted by self-regard (Herrmann, Thoni, and Gächter 2008). In both JDG and PDG, we find that greater prevalence of destructive and defective behavior holds even when subjects make decisions when the counterpart was non-destructive/cooperative, and thus when there is no scope for perceiving the counterpart's actions as hostile. In fact, the magnitude of the effects of the group context is often somewhat larger when the counterpart was not destructive, as compared to when the counterpart was destructive, indicating that a group context influences behavior especially in situations in which strong social norms against destruction exist. Thus, our results reveal that a group context leads to greater prevalence of pure destructiveness rather than to a stronger preference to retaliate hostile behavior.

E.3. Mechanisms

In this sub-section, we discuss why individuals become more destructive when deciding on behalf of a group and consider which of the following potential mechanisms -- diffused responsibility, social image considerations, pleasing other group members, or ingroup/outgroup biases -- can best explain the patterns we find.

The observed increase in destructiveness in *GC_Hidden*, compared to *Individual*, provides support for the interpretation that the higher prevalence of destructiveness is primarily driven by reduced moral costs to one's self-image from behaving anti-socially. 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 perception of individual responsibility for the decision, as compared to decision-making in *Individual*, in which it is salient that the decision-maker is fully responsible for the ultimate action. At the same time, there is very little scope for several alternative features of the group context that may also affect social behavior, such as observability and salient group boundary. Yet, we still find systematic effects.

The increased destructiveness in the group context is unlikely to be due to social image considerations or a desire of some subjects to signal to others that they are "tough" types for strategic reasons, e.g. due to expectations of future repeated interactions outside of the lab. For example, subjects may fear anti-social members of their group, and thus may choose the destructive strategy to avoid the risk of becoming an object of ridicule for being too "weak". This mechanism predicts that there should be an increase in the prevalence of destructive behavior in *GC_Observed*, as compared to *GC_Hidden*, because in *GC_Observed* the subjects knew the other group members' identity and that they would observe each other's decisions, in contrast to *GC_Hidden*, in which choices were made in private and subjects did not know the matched group members. However, we do not observe such a pattern. In fact, we observe that the effects are somewhat attenuated in *GC_Observed*.[20]

[20] Similarly, this pattern also does not favor the interpretation that the increase in destructive behavior is due to utility from a *shared* experience of winning and beating outsiders, because this mechanism is relevant only when subjects know each other and can observe each other (i.e., in *GC_Observed* but not in *GC_Hidden*)

Next, we consider the possibility that subjects in a group context may aim to please other group members by acting destructively, out of intrinsic motivations. For this mechanism to be relevant, it is necessary that subjects believe that the other group members prefer the destructive choice to the non-destructive choice. In order to gauge the relevance of this mechanism, we have elicited beliefs about the unconditional decisions made by the other two group members. Figure A.2 reports the distribution of these beliefs. We find that a much larger percentage of subjects (46%) in *GC_Hidden* expect both group members to be non-destructive as 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.

Finally, a potential mechanism is that, when individuals are allocated to salient groups, they develop an "ingroup vs. outgroup" psychology, 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, as compared to *Individual*, we observe an increase in destructiveness in *GC_Observed*, a condition in which group boundaries are salient because group members personally interacted with each other. However, importantly, the effects arise (and are even stronger) in *GC_Hidden*, in which group boundary is not salient. In *GC_Hidden*, subjects were matched into a group with two other anonymous group members. No group attribute was shared by the members, which would distinguish their group from the experimental counterpart group. In fact, the information provided about the members of their own group and the members of the counterpart group was identical – they were anonymous university students participating in the research in another session. Thus, unless the reference to a "group" alone is enough to trigger ingroup/outgroup biases, there was no scope for this mechanism to play a role in *GC_Hidden*.[21]

[21] In Experiments 1 and 2, another mechanism through which a group context could have affected decisions was the expectation of group deliberation, since in *GC_1* subjects were aware that joint decision-making would follow. In Experiment 3, in contrast, this mechanism is closed down by design: the group decision-making came as a surprise, after subjects completed their decisions in the *Individual*, *GC_Hidden* or *GC_Observed* conditions.

E.4 Effects of group decision-making

In this sub-section, we estimate the effects of group decision-making. To do so, we compare behavior in the *Group* condition and in the group context conditions. We find a clear pattern: group decision-making increases the prevalence of behavior that maximizes the payoff of the whole group (Figure 4, Table 3).

In JDG, when looking at the index constructed from all three choices, the prevalence of destructive behavior diminishes from 27% in *GC_Hidden* and 24% in *GC_Observed* to 16% in *Group* (p-value<0.01 for both comparisons). We observe similar effects for the unconditional choice and for both conditional choices (Panel A of Table 3). Thus, we find that in this task, group decision-making has the opposite effects from the effects of the group context. Given that the prevalence of destructive choices in *Individual* is 20%, group deliberation and joint decision-making eliminates (and for the unconditional decision more than compensates) the effects of the group context on higher prevalence of destructive behavior.

In PDG, the defection rate increases from 64% in both *GC_Hidden* and *GC_Observed* to 72% in *Group* (p-value<0.01 for both comparisons). Again, the effects are similar for all three choices in this task (Panel B of Table 3). Thus, in PDG, joint decision making reinforces the effects of the group context - both effects contribute to greater defection of groups than individuals. Since group decision-making decreases destructiveness in JDG and reduces cooperativeness in PDG, letting university students to make decisions jointly as a group, rather than making individual choices on behalf of the group, increases the likelihood of playing strategies predicted for agents with purely self-regarding preferences.

Next, we consider two distinct mechanisms through which group decision-making may affect choices. The first, which we refer to as preference-aggregation, is the outcome of the fact that many groups agree with the initial opinion of the median member, and thus joint decisions naturally reduce the prevalence of minority opinions. The second is the effects of group deliberation, which involves the role of group discussion, information exchange and persuasion, and which may motivate the groups to depart from the behavior predicted by the initial judgement of the median group member.

We first consider whether the observed reduction in destructiveness and cooperativeness in *Group* can be fully explained by members agreeing on the action a priori preferred by a median member of the group. To do so, for each group we simulate a joint decision by assuming that it is fully determined by the choice of the median group member (Table A.9). Since in JDG the destructive option was chosen by a minority of subjects (in the unconditional decision and the choice conditional on a counterpart being non-destructive), the simulated decisions imply a substantial reduction in destructiveness. Nevertheless, we find that such aggregation of preferences cannot fully explain the observed effect of group decision-making. We see further reduction in destructiveness in actual group decisions, beyond the simulated decisions, suggesting that group deliberation also contributes to reduced destructiveness among university students.

We further study the role of group deliberation by exploring how the composition of preferences among individual group members, based on their choices in *Individual*, *GC_Hidden* and *GC_Observed,* is predictive of actual group decisions (Table A.10). Overall, we find that most groups (97%) follow a median a priori opinion when such an opinion implies maximization of their own group payoffs, and thus only a relatively low percentage of groups (3%) switches to the destructive strategy in JDG after joint discussion and deliberation (Panel A of Table A.10). In contrast, when the initial judgement of the median group member is to destroy (there are relatively few such cases), a fairly large percentage of the groups (67%) ultimately agree on the non-destructive strategy. Thus, among university students the idea to maximize their own group's payoff is more persuasive during group deliberation than the possibility to destroy the earnings of the counterpart.

In PDG, we find a similar pattern - greater prevalence of switching to payoff maximizing behavior after the group deliberation (Panel B of Table A.10). Specifically, more than twice as many groups switch to defection after group deliberation (when the initial median opinion was cooperation), as compared to the percentage of groups which switch to cooperation (when the initial median opinion predicted defection).[22]

[22] In the Online Appendix, we perform similar analyses for Experiments 1 and 2 (Tables A.17 and A.18). In PDG, we find similar patterns as among university students – greater prevalence of switching to defection from behavior implied by a median member, rather than vice versa, suggesting that group deliberation facilitates a shift towards defection in PDG also among adolescents. In JDG, we do not find such asymmetric switching from the initial median opinion. In contrast to adults, group deliberation does not facilitate a shift towards non-destructive behavior among adolescents. This helps to explain why there is virtually no effect of group decision-making in JDG for the younger sample.

E.5 Summary

Experiment 3 provides the following insights. First, a group context also marks individual destructive behavior among university students. When they make decisions on behalf of a group rather than individually, they choose to behave more destructively in JDG and less cooperatively in PDG. Thus, we replicate the main pattern observed among adolescents in the earlier Experiments 1 and 2. Further, we test for several alternative mechanisms and the results favor the interpretation that the "destructiveness shift" is primarily driven by a reduced perception of individual responsibility when acting on behalf of a group.

Second, the group decision-making process is also influential, but differs qualitatively from the effects of a group context. 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 individuals act alone or individuals act in a group context. We find that this effect is more systematic among university students, since group decision-making increases selfishness at the expense of both cooperative and destructive behaviors. Among the adolescents in Experiments 1 and 2, group decision-making increases the prevalence of selfish behavior only due to the reduced prevalence of cooperative behavior; destructive behavior was not reduced by group decision-making.

III. EXPERIMENT 4: ROBUSTNESS AMONG A NATIONALLY REPRESENTATIVE SAMPLE OF ADULTS

In the second follow-up study, we focus on the main finding from the previous studies: when subjects make decisions on behalf of a group rather than individually, their behavior becomes nastier. In Experiment 3, we show that this effect holds not only among adolescents but also among a sample of university students. However, university students are a narrow sub-set of the adult population. In order to test whether the effect holds for adults across various demographic and economic groups, in Experiment 4 we study how decision-making on behalf of a group affects destructiveness among a large nationally representative sample of adults in the same country.

A. Sample

Experiment 4 was implemented among a sample of 4,243 adults in the Slovak Republic. The data collection was conducted online in cooperation with a survey company, IPSOS. The sample consists of participants 18 years of age or older who were quota sampled from an actively recruited online panel of the survey company in order to be balanced and representative of the general population based on a set of observable characteristics, in terms of 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 (Table A.11). Distributions of other characteristics are similar.[23]

B. Experimental tasks

As in Experiments 1-3, the subjects participated in the Joy of Destruction mini-game (JDG) and in the Prisoner's Dilemma game (PDG), in randomized order. In JDG, both players received EUR 2 and made a decision whether to pay EUR 0.2 to destroy EUR 1 belonging to the counterpart, or whether to keep the payoffs unchanged (Figure 1). In PDG, both players received EUR 1.6 and decided whether to take EUR 0.8 of the counterpart to gain EUR 0.4 for themselves, or whether to keep the payoffs unchanged. The payoff matrices are thus identical to those used in Experiment 1 with adolescents.

In each game, participants made an unconditional decision, stated their beliefs regarding the choice of their counterpart, and 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 (as in Experiment 3). All choices were incentivized. The participants were paid for one randomly selected decision.

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 and change the status quo when acting on behalf of a group. In both

[23] Experiment 4 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. Participation in the experiment was voluntary and the subjects could stop responding to the questions at any time.

JDG and PDG, the choice was either to keep the payoffs unchanged or to take an action and reduce part of the counterpart's payoff. Thus, if "action bias" was stronger when individuals acted on behalf of a group, it could explain the pattern we find. 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 had by default either yellow or blue color. The respondents could choose to keep the color (opt for a passive option) or to change it (choose an active option). Depending on the condition, subjects made this decision individually or on behalf of a group.

C. Experimental conditions

In Experiment 4, we implemented only the *Individual* and *GC_Hidden* conditions. This design choice was guided by the following considerations. First, Experiment 3 showed that, in order to identify the "destructiveness shift" in groups, it is sufficient to let people act on behalf of a completely meaningless, anonymous group. Second, this condition has the most straightforward interpretation in terms of its links to the underlying mechanism. The third aspect was feasibility; it was not possible to implement *GC_Observed* and *Group* conditions using an online survey infrastructure, as they require face-to-face, real-time interactions among group members.

The subjects were allocated to either *Individual* or *GC_Hidden* condition, using a between-subject design. The *GC_Hidden* condition is very similar to the *GC_Hidden* condition implemented among university students in Experiment 3. 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 a 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 on behalf of their group. Online Appendix B contains the full experimental protocol.

Since in Experiment 4 we have such a diverse sample, we gathered a very rich set of observable participant characteristics. In Table A.11 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 separately followed a detailed quota sampling procedure, in order 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 Experiments 1-3.

D. Results

D.1 Effects of group context

We find that the group context has an important influence on behavior (Figure 5 and Table 4), replicating the findings of Experiments 1, 2 and 3. We first analyze the index of destructive behavior, constructed as the average of three indicator variables of a destructive choice (unconditional decision and two conditional choices). In JDG (Panel A of Table 4), participants in *Individual* chose the destructive option in 21% of cases, while participants in *GC_Hidden* chose the destructive option in 29% of cases. Thus, making a decision on behalf of an anonymous group rather than on one's own behalf increases the prevalence of destructive behavior by around 37% (p-value < 0.001, Fig. 5, Table 4). The results are robust to controlling for the order of the games and hold when we control for a range of observable characteristics (Table A.12).

We observe similar effects for each of the three choices in JDG (Panel A of Table 4). In *GC_Hidden*, the prevalence of destructive choices increases by 7 p.p. in the unconditional decision (from 18.5% in *Individual*, p-value < 0.001), by 8 p.p. when the counterpart is destructive (from 32%, p-value < 0.001), and by 8 p.p. when the counterpart is not destructive (from 13%, p-value < 0.001). As in Experiment 3, these patterns reveal that greater prevalence of destructiveness in

the anonymous group context 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. The effect is largest (61% increase) when subjects choose whether to destroy the resources of a counterpart who decided to act non-destructively, and thus in a decision situation in which it is most socially unacceptable to act in a hostile manner. Thus, the results are consistent with the interpretation that the group context reduces the scruples against deviating from social norms and giving in to intrinsic urges to destroy.

An anonymous group context also influences behavior in PDG (Panel B of Table 4). When focusing on the index, the defection rate rises from 35% in *Individual* to 42% in *GC_Hidden* (p-value < 0.001). Similar effects are observed for each of the three decisions. The defection rate increases by 7 p.p. in the unconditional decision (from 24% in *Individual*, p-value < 0.001), by 5 p.p. when the counterpart is not cooperative (from 56%, p-value = 0.003), and by 11 p.p. when the counterpart is cooperative (from 23%, p-value < 0.001). Again, the magnitude of the effect is the most profound in relative terms (46% increase) when defection is least socially acceptable, i.e. when the counterpart is cooperative.

Further, we show that the effects of *GC_Hidden* are unlikely to be driven by a greater tendency to choose an active option rather that to passively stick with the default. Specifically, in the "action-bias" task, participants in *GC_Hidden* are not more likely to change the color of an emoticon from its default color than those allocated to *Individual* (Table A.13).[24]

D.2 Heterogeneity

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 for 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, 65 and up. In terms of education, we distinguish three groups based on the highest completed education level: primary or lower secondary, upper

[24] We arrive at the same conclusions in Experiment 3, where we measured the „action bias" using a different task in *Individual* and *Group* conditions. The Online Appendix provides more details.

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, more than 100. Finally, we classify participants based on their political party preferences across two dimensions: economic right/left wing (right, middle, left) and liberal/conservative (liberal, neutral, conservative).[25]

The effects are robust. We find that the anonymous group context increases destructive behavior among all the demographic sub-groups we study (Figure 6 and Table A.14). The effects are somewhat larger for males than for females, and for individuals between 25-44 than for other age groups, but remain statistically significant for all sub-groups (except those who are 65 or older). The effects on behavior in PDG are similarly robust (Figure A.3 and Table A.15).

IV. 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 are involved in making decisions on behalf of a group rather than when making their own individual decisions. We establish this behavioral regularity across a broad range of demographic and socio-economic groups, by conducting four large-scale experiments among adolescents, university students and a nationally representative sample of adults -- more than seven thousand subjects in total. We test several potential mechanisms; the results favor the interpretation that the behavioral change is driven mainly by a lessened perception of responsibility when individuals make decisions in the group context.

The findings have several implications. First, they provide empirical support for the idea that perception of individual responsibility is an important regulator of the dark side of human personality, in line with self-signaling models. They illuminate that, in the 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, the results help to explain why decisions made by groups tend to be less pro-social than individual decisions. We show that this is partly due to the effect of group context on nastiness, and partly

[25] We provide more details about the classification of political parties in Table A.14.

due to the role of group deliberation, which fosters a "selfishness shift", a mechanism featured in earlier work (Charness and Sutter 2012; Kugler, Kausel, and Kocher 2012). Third, in terms of practical implications, our findings suggest that organizations or public bodies seeking to limit obstructionism and other manifestations of nasty inclinations, may want to create environments that foster perception of individual responsibility. For example, by framing decisions as the choices of specific individuals, rather than presenting them as decisions of the whole units, teams or committees, or by making the ways actions of individual group members translate into group outcomes more transparent.

Finally, inspired by evidence showing that people tend to seek environments where 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 context effect we identify. If people are aware that an anonymous group context constitutes an environment in which 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 this could be a fruitful avenue to explore.

REFERENCES

Abbink, Klaus, and Benedikt Herrmann. 2011. "The Moral Costs of Nastiness." *Economic Inquiry* 49 (2): 631–33.

Abbink, Klaus, and Abdolkarim Sadrieh. 2009. "The Pleasure of Being Nasty." *Economics Letters* 105 (3): 306–8.

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

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 (5): 1607–1636.

Balafoutas, Loukas, Rudolf Kerschbamer, Martin Kocher, and Matthias Sutter. 2014. "Revealed Distributional Preferences: Individuals vs. Teams." *Journal of Economic Behavior and Organization* 108: 319–30.

Bauer, Michal, Jana Cahlíková, Julie Chytilová, and Tomáš Želinský. 2018. "Social Contagion of Ethnic Hostility." *Proceedings of the National Academy of Sciences* 115 (19): 4881–86.

Bauer, Michal, Julie Chytilová, and Edward Miguel. 2020. "Using Survey Questions to Measure Preferences : Lessons from an Experimental Validation in Kenya." *European Economic Review* 127:

103493.
Bauer, Michal, Julie Chytilová, and Barbara Pertold-Gebicka. 2014. “Parental Background and Other-Regarding Preferences in Children.” *Experimental Economics* 17 (1): 24–46.
Benabou, Roland, Armin Falk, and Jean Tirole. 2018. “Narratives, Imperatives, and Moral Reasoning.” NBER Working Paper No. 24798.
Benabou, Roland, and Jean Tirole. 2006. “Incentives and Prosocial Behavior.” *American Economic Review* 96 (5): 1652–78.
———. 2011. “Identity, Moral and Taboos: Beliefs as Assets.” *Quarterly Journal of Economics* 126: 805–55.
Bon, Gustave Le. 1895. *The Crowd*. New York: Viking (1960).
Bornstein, Gary, and Ilan Yaniv. 1998. “Individual and Group Behavior in the Ultimatum Game: Are Groups More " Rational " Players?” *Experimental Economics* 1 (1): 101–8.
Bursztyn, Leonardo, and Robert Jensen. 2017. “Social Image and Economic Behavior in the Field : Identifying , Understanding and Shaping Social Pressure.” *Annual Review of Economics* 9 (1): 131–53.
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): 157–76.
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, 197–216. New York: Routledge.
Dana, Jason, Roberto Weber, and Jason Xi Kuang. 2006. “Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness.” *Economic Theory* 33 (1): 67–80.
Darley, John M., and Bibb Latane. 1968. “Bystander Intervention in Emergencies: Diffusion of Responsibility.” *Journal of Personality and Social Psychology* 8 (4): 377–83.
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): 1–56.
Deming, David J. 2011. “Better Schools, Less Crime?” *The Quarterly Journal of Economics* 126 (4): 2063–2115.
Diener, Edward. 1980. “Deindividuation: The Absence of Self-Awareness and Self-Regulation in Group Members.” In *The Psychology of Group Influence*, edited by P.B. Paulus, 209–242. Hills-dale, NJ: Erlbaum.
Durlauf, Steven N. 1999. “The Case ‘against’ Social Capital.” *FOCUS Newsletter for the Institute for Research on Poverty* 20 (3): 1–5.
Falk, Armin. 2017. “Facing Yourself : A Note on Self-Image.” IZA Discussion Paper 10606.
Falk, Armin, Ernst Fehr, and Urs Fischbacher. 2005. “Driving Forces behind Informal Sanctions.” *Econometrica* 73 (6): 2017–2030.
Falk, Armin, Thomas Neuber, and Nora Szech. 2020. “Diffusion of Being Pivotal and Immoral Outcomes.” *Review of Economic Studies* 87 (5): 2205–29.
Fehr, Ernst, and Simon Gächter. 2002. “Altruistic Punishment in Humans.” *Nature* 415 (6868): 137–40.
Fehr, Ernst, Karla Hoff, and Mayursh Kshetramade. 2008. “Spite and Development.” *American Economic Review: Papers and Proceedings* 98 (2): 494–499.
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.
Gächter, Simon, Elke Renner, and Martin Sefton. 2008. “The Long-Run Benefits of Punishment.” *Science* 322 (5907): 1510–1510.
Gintis, Herbert, Samuel Bowles, Robert T. Boyd, and Ernst Fehr. 2005. *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*. MIT Press.

Herrmann, Benedikt, Christian Thoni, and Simon Gächter. 2008. "Antisocial Punishment across Societies." *Science* 319 (5868): 1362–67.
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 (1): 228–38.
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 (6): 646–57.
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 (4): 471–82.
Lochner, Lance, and Enrico Moretti. 2004. "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports." *American Economic Review* 94 (1): 155–89.
Luhan, Wolfgang J., Martin G. Kocher, and Matthias Sutter. 2009. "Group Polarization in the Team Dictator Game Reconsidered." *Experimental Economics* 12 (1): 26–41.
Maniadis, Zacharias, Fabio Tufano, and John A. List. 2015. "How to Make Experimental Economics Research More Reproducible: Lessons from Other Disciplines and a New Proposal." In *Replication in Experimental Economics*, edited by C. Deck, E. . Fatas, and T. Rosenblat, 215–30. Emerald Group Publishing Limited.
Murphy, Ryan O., and Kurt A. Ackermann. 2014. "Social Value Orientation : Theoretical and Measurement Issues in the Study of Social Preferences." *Personality and Social Psychology Review* 18 (1): 13–41.
Neuman, Joel H., and Robert A. Baron. 1998. "Workplace Violence and Workplace Aggression : Evidence Concerning Specific Forms , Potentiai Causes , and Preferred Targets." *Journal of Management* 24 (3): 391–419.
Niederle, Muriel, and Lise Vesterlund. 2007. "Do Women Shy Away From Competition? Do Men Compete Too Much?" *The Quarterly Journal of Economics* 122 (3): 1067–1101.
Prediger, Sebastian, Bjorn Vollan, and Benedikt Herrman. 2014. "Resource Scarcity and Antisocial Behavior." *Journal of Public Economics* 119: 1–9.
Publius. 1948. "Federalist No. 55." In *The Federalist*, edited by M. Beloff, 283–287. Oxford, England: Blackwell (Original work published 1788).
Schopler, John, and Chester A. Insko. 1992. "The Discontinuity Effect in Interpersonal and Intergroup Relations: Generality and Mediation." *European Review of Social Psychology* 3 (1): 121–51.
Sutter, Matthias. 2009. "Individual Behavior and Group Membership: Comment." *American Economic Review* 99 (5): 2247–57.
Tajfel, Henry. 1981. *Human Groups and Social Categories: Studies in Social Psychology*. Cambridge: Cambridge University Press.
Tajfel, Henry, and John Turner. 1979. "An Integrative Theory of Intergroup Conflict." In *The Social Psychology of Intergroup Relations*, edited by Stephen Worchel and William Austin. Monterey, CA: Brooks/Cole.
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 (5): 698–722.

TABLE 1— PROPORTION OF DESTRUCTIVE AND NON-COOPERATIVE CHOICES, EXPERIMENT 1 (SLOVAKIA, ADOLESCENTS)

	Means				Effect of group context		Effect of group decision-making		Overall effect
	Individual	Group Context_1	Group Context_2	Group	Diff (2)-(1) [p-value]	Diff (3)-(1) [p-value]	Diff (4)-(2) [p-value]	Diff (4)-(3) [p-value]	Diff (4)-(1) [p-value]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Joy of Destruction game, proportion of destructive choices									
Index (0=min, 1=max)	0.39	n/a	0.47	0.44	n/a	8 [0.039]	n/a	-3 [0.341]	5 [0.246]
Unconditional decision	0.32	0.44	0.45	0.42	12 [0.004]	13 [0.026]	-3 [0.482]	-3 [0.546]	10 [0.055]
Conditional decision, if partner non-destructive	0.29	n/a	0.42	0.38	n/a	12 [0.031]	n/a	-3 [0.541]	9 [0.057]
Conditional decision, if partner destructive	0.55	n/a	0.54	0.51	n/a	-1 [0.933]	n/a	-4 [0.539]	-4 [0.432]
Beliefs: Counterpart destructive	0.43	n/a	0.56	0.42	n/a	13 [0.021]	n/a	-13 [0.015]	0 [0.948]
N	222	1,021	158	346	1,243	380	1,367	504	568
Panel B: Prisoner's Dilemma game, proportion of non-cooperative choices									
Index (0=min, 1=max)	0.74	n/a	0.79	0.84	n/a	5 [0.252]	n/a	6 [0.057]	10 [<0.001]
Unconditional decision	0.67	0.75	0.72	0.82	8 [0.035]	4 [0.436]	7 [0.014]	11 [0.024]	15 [<0.001]
Conditional decision, if partner cooperative	0.68	n/a	0.78	0.80	n/a	9 [0.061]	n/a	2 [0.618]	12 [0.005]
Conditional decision, if partner non-cooperative	0.87	n/a	0.87	0.91	n/a	0 [0.912]	n/a	4 [0.275]	4 [0.179]
Beliefs: Counterpart non-cooperative	0.66	n/a	0.70	0.73	n/a	4 [0.456]	n/a	3 [0.493]	7 [0.118]
N	222	1,036	158	346	1,258	380	1,382	504	568

Notes: Panel A reports the prevalence of destructive choices and beliefs that the counterpart was destructive in the Joy of Destruction game, while Panel B reports the prevalence of non-cooperative choices and beliefs that the counterpart was non-cooperative in the Prisoner's Dilemma game. "Index" is calculated as an average of three indicator variables for whether a decision-maker acted destructively/non-cooperatively in the unconditional choice and in two conditional choices. Columns 1 – 4 report choices and beliefs separately for the four experimental conditions. Columns 5 – 9 present differences between the conditions in percentage points. In brackets, we show the associated Somers' D p-values adjusted for clustering at individual level, as we have two observations per subject (a choice regarding a Roma counterpart and a choice regarding a majority-ethnicity counterpart).

TABLE 2— PROPORTION OF DESTRUCTIVE AND NON-COOPERATIVE CHOICES, EXPERIMENT 2 (UGANDA, ADOLESCENTS)

	Means			Effect of group context	Effect of group decision-making	Overall effect
	Individual	Group Context_1	Group	Diff (2)-(1) [p-value]	Diff (3)-(2) [p-value]	Diff (3)-(1) [p-value]
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Joy of Destruction game, proportion of destructive choices						
Index (0=min, 1=max)	0.59	n/a	0.65	n/a	n/a	6 [0.002]
Unconditional decision	0.53	0.60	0.59	7 [0.012]	0 [0.910]	7 [0.052]
Conditional decision, if partner non-destructive	0.44	n/a	0.52	n/a	n/a	8 [0.020]
Conditional decision, if partner destructive	0.80	n/a	0.84	n/a	n/a	5 [0.084]
Beliefs: Counterpart destructive	0.60	n/a	0.62	n/a	n/a	2 [0.615]
N	427	1,236	417	1,663	1,653	844
Panel B: Prisoner's Dilemma game, proportion of non-cooperative choices						
Index (0=min, 1=max)	0.59	n/a	0.74	n/a	n/a	15 [<0.001]
Unconditional decision	0.57	0.68	0.79	10 [<0.001]	11 [<0.001]	22 [<0.001]
Conditional decision, if partner cooperative	0.39	n/a	0.54	n/a	n/a	15 [<0.001]
Conditional decision, if partner non-cooperative	0.80	n/a	0.87	n/a	n/a	8 [0.003]
Beliefs: Counterpart non-cooperative	0.66	n/a	0.68	n/a	n/a	2 [0.495]
N	428	1,238	417	1,666	1,655	845

Notes: Panel A reports the prevalence of destructive choices and beliefs that the counterpart was destructive in the Joy of Destruction game, while Panel B reports the prevalence of non-cooperative choices and beliefs that the counterpart was non-cooperative in the Prisoner's Dilemma game. "Index" is calculated as an average of three indicator variables for whether a decision-maker acted destructively/non-cooperatively in the unconditional choice and in two conditional choices. Columns 1 – 3 report choices and beliefs separately for the three experimental conditions. Columns 4 – 6 present differences between the conditions in percentage points. In brackets, we show the associated p-values from a Wilcoxon rank-sum test (Index) or a Chi-square test (binary outcomes).

TABLE 3— PROPORTION OF DESTRUCTIVE AND NON-COOPERATIVE CHOICES, EXPERIMENT 3 (SLOVAKIA, UNIVERSITY STUDENTS)

	Means				Effect of group context		Effect of group decision-making		Overall effect
	Individual	Group Context_Hidden	Group Context_Observed	Group	Diff (2)-(1) [p-value]	Diff (3)-(1) [p-value]	Diff (4)-(2) [p-value]	Diff (4)-(3) [p-value]	Diff (4)-(1) [p-value]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Joy of Destruction game, proportion of destructive choices									
Index (0=min, 1=max)	0.20	0.27	0.24	0.16	7 [0.001]	4 [0.051]	-11 [<0.001]	-8 [0.001]	-3 [0.164]
Unconditional decision	0.12	0.18	0.12	0.05	5 [0.090]	-1 [0.771]	-12 [<0.001]	-6 [0.009]	-7 [0.004]
Conditional decision, if partner non-destructive	0.04	0.13	0.07	0.02	10 [<0.001]	3 [0.143]	-12 [<0.001]	-5 [0.015]	-2 [0.297]
Conditional decision, if partner destructive	0.43	0.51	0.54	0.42	7 [0.083]	10 [0.019]	-9 [0.044]	-12 [0.009]	-1 [0.761]
Beliefs: Counterpart destructive	0.15	0.24	0.15	0.09	9 [0.009]	0 [0.891]	-15 [<0.001]	-6 [0.037]	-6 [0.026]
N	267	267	261	254	534	528	521	515	521
Panel B: Prisoner's Dilemma game, proportion of non-cooperative choices									
Index (0=min, 1=max)	0.59	0.64	0.64	0.72	5 [0.068]	5 [0.082]	7 [0.007]	7 [0.006]	12 [<0.001]
Unconditional decision	0.52	0.53	0.51	0.61	0 [0.931]	-1 [0.734]	9 [0.047]	10 [0.017]	9 [0.039]
Conditional decision, if partner cooperative	0.33	0.46	0.46	0.55	13 [0.002]	13 [0.002]	9 [0.039]	9 [0.047]	22 [<0.001]
Conditional decision, if partner non-cooperative	0.93	0.93	0.95	0.98	0 [0.865]	2 [0.304]	5 [0.010]	3 [0.091]	5 [0.006]
Beliefs: Counterpart non-cooperative	0.54	0.56	0.56	0.65	3 [0.543]	2 [0.645]	9 [0.032]	10 [0.023]	12 [0.006]
N	267	267	261	254	534	528	521	515	521

Notes: Panel A reports the prevalence of destructive choices and beliefs that the counterpart was destructive in the Joy of Destruction game, while Panel B reports the prevalence of non-cooperative choices and beliefs that the counterpart was non-cooperative in the Prisoner's Dilemma game. "Index" is calculated as an average of three indicator variables for whether a decision-maker acted destructively/non-cooperatively in the unconditional choice and in two conditional choices. Columns 1 – 4 report choices and beliefs separately for the four experimental conditions. Columns 5 – 9 present differences between the conditions in percentage points. In brackets, we show the associated p-values from a Wilcoxon rank-sum test (Index) or a Chi-square test (binary outcomes).

TABLE 4— PROPORTION OF DESTRUCTIVE AND NON-COOPERATIVE CHOICES, EXPERIMENT 4 (SLOVAKIA, REPRESENTATIVE SAMPLE OF ADULT POPULATION)

	Means		Effect of group context
	Individual	Group Context _Hidden	Diff (2)-(1) [p-value]
	(1)	(2)	(3)
Panel A: Joy of Destruction game, proportion of destructive choices			
Index (0=min, 1=max)	0.21	0.29	8 [<0.001]
Unconditional decision	0.19	0.26	7 [<0.001]
Conditional decision, if partner non-destructive	0.13	0.21	8 [<0.001]
Conditional decision, if partner destructive	0.32	0.41	8 [<0.001]
Beliefs: Counterpart destructive	0.30	0.36	6 [<0.001]
N	2141	2102	4,243
Panel B: Prisoner's Dilemma game, proportion of non-cooperative choices			
Index (0=min, 1=max)	0.35	0.42	7 [<0.001]
Unconditional decision	0.24	0.31	7 [<0.001]
Conditional decision, if partner cooperative	0.23	0.34	11 [<0.001]
Conditional decision, if partner non-cooperative	0.56	0.61	5 [0.003]
Beliefs: Counterpart non-cooperative	0.38	0.42	5 [0.001]
N	2141	2102	4,243

Notes: Panel A reports the prevalence of destructive choices and beliefs that the counterpart was destructive in the Joy of Destruction game, while Panel B reports the prevalence of non-cooperative choices and beliefs that the counterpart was non-cooperative in the Prisoner's Dilemma game. "Index" is calculated as an average of three indicator variables for whether a decision-maker acted destructively/non-cooperatively in the unconditional choice and in two conditional choices. Columns 1 – 2 report choices and beliefs separately for the two experimental conditions. Column 3 presents differences between the conditions in percentage points. In brackets, we show the associated p-values from a Wilcoxon rank-sum test (Index) or a Chi-square test (binary outcomes).

FIGURE 1. PAYOFF MATRICES

PANEL A: Experiment 1 (Slovakia, adolescents); in EUR

a) Joy of Destruction game

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

b) Prisoner's Dilemma game

		Player B:	
		Cooperate	Defect
Player A:	Cooperate	1.6, 1.6	0.8, 2
	Defect	2, 0.8	1.2, 1.2

PANEL B: Experiment 2 (Uganda, adolescents); in USh

a) Joy of Destruction game

		Player B:	
		Non-destructive	Destructive
Player A:	Non-destructive	1000, 1000	500, 900
	Destructive	900, 500	400,400

b) Prisoner's Dilemma game

		Player B:	
		Cooperate	Defect
Player A:	Cooperate	800, 800	400, 1000
	Defect	1000, 400	600, 600

PANEL C: Experiment 3 (Slovakia, university students); in EUR

a) Joy of Destruction game

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

b) Prisoner's Dilemma game

		Player B:	
		Cooperate	Defect
Player A:	Cooperate	8, 8	4, 10
	Defect	10, 4	6, 6

PANEL D: Experiment 4 (Slovakia, representative sample of adult population); in EUR

a) Joy of Destruction game

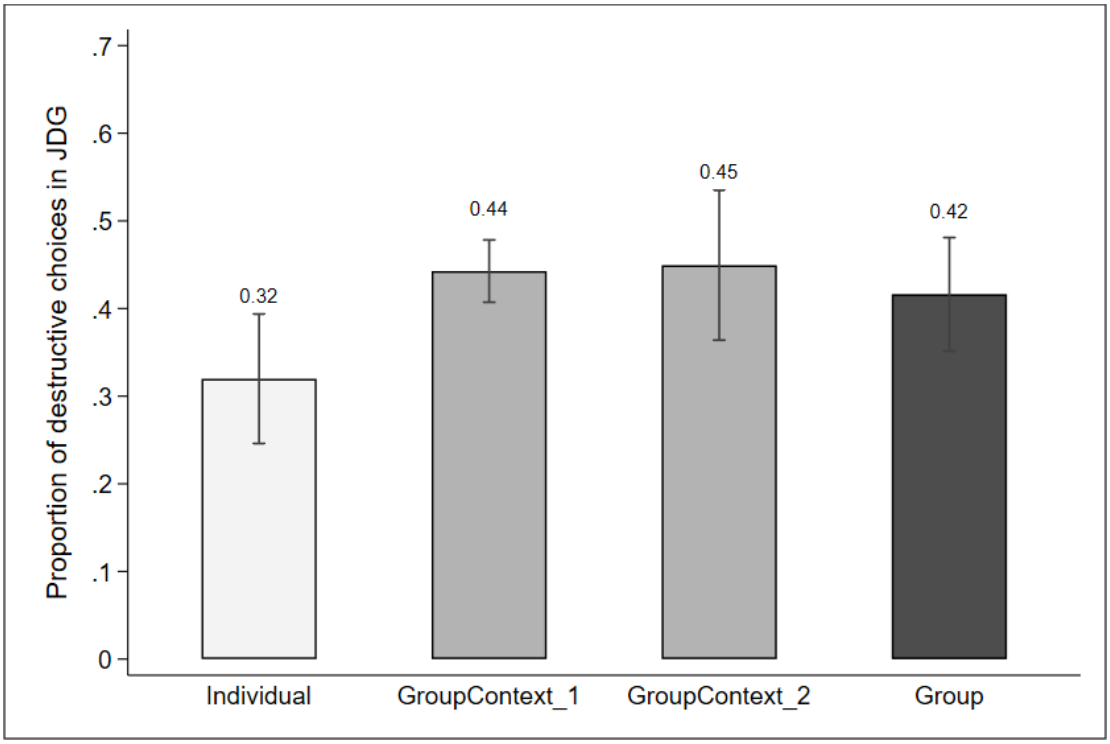
		Player B:	
		Non-destructive	Destructive
Player A:	Non-destructive	2, 2	1.8, 1
	Destructive	1.8, 1	0.8, 0.8

b) Prisoner's Dilemma game

		Player B:	
		Cooperate	Defect
Player A:	Cooperate	1.6, 1.6	0.8, 2
	Defect	2, 0.8	1.2, 1.2

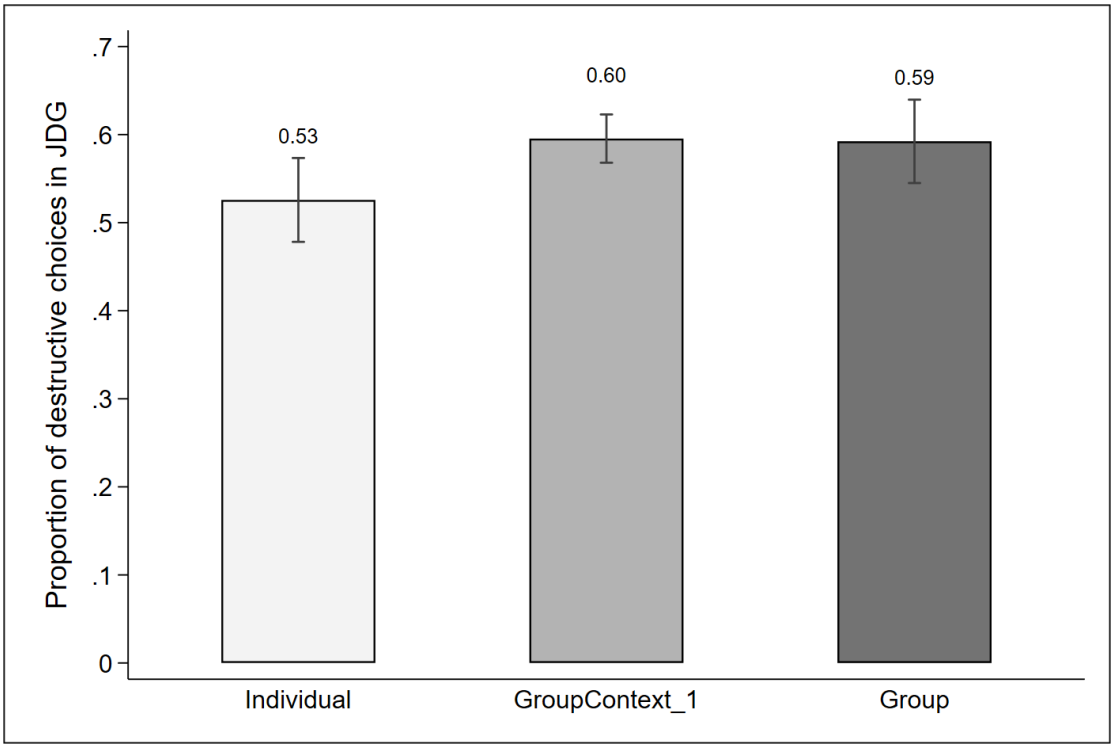
Notes: Payoff matrices for games implemented in the experiments. The first number in the matrix cell indicates the payoff of Player A and the second number indicates the payoff of Player B.

FIGURE 2. DESTRUCTIVE BEHAVIOR IN GROUPS, EXPERIMENT 1 (SLOVAKIA, ADOLESCENTS)



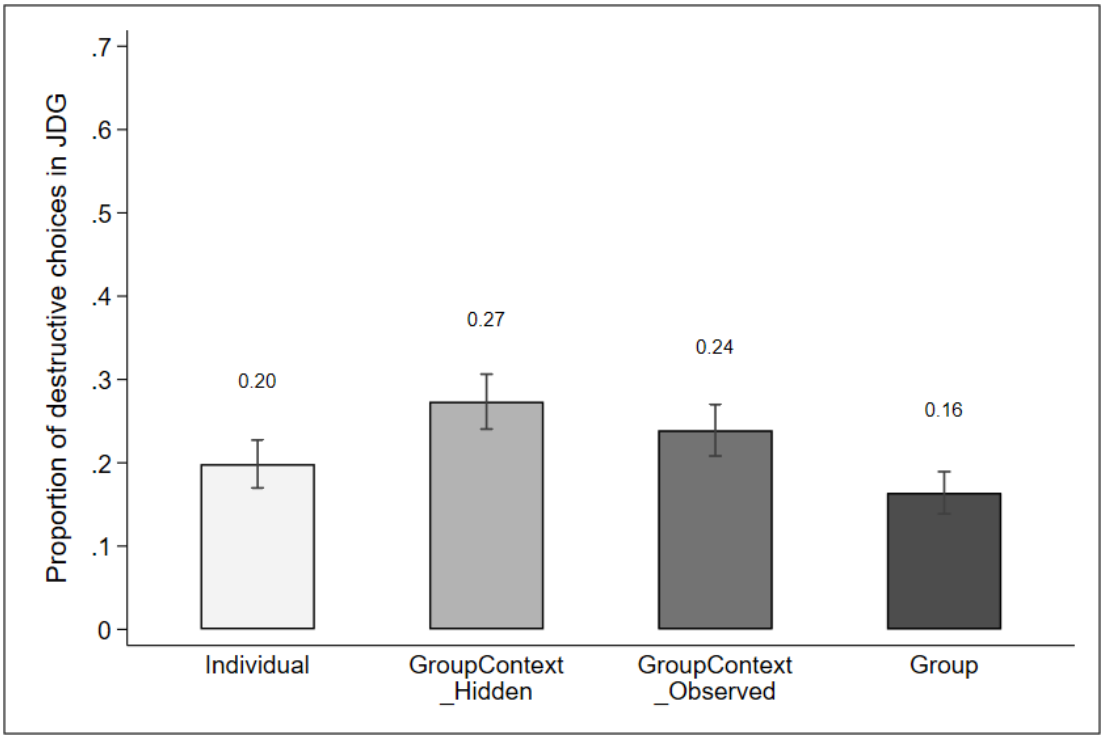
Notes: The proportion of unconditional destructive choices in the Joy of Destruction game. Bars indicate 95% confidence intervals. Standard errors are clustered at an individual level, as we have two observations per subject (a choice regarding a Roma counterpart and a choice regarding a majority-ethnicity counterpart).

FIGURE 3. DESTRUCTIVE BEHAVIOR IN GROUPS, EXPERIMENT 2 (UGANDA, ADOLESCENTS)



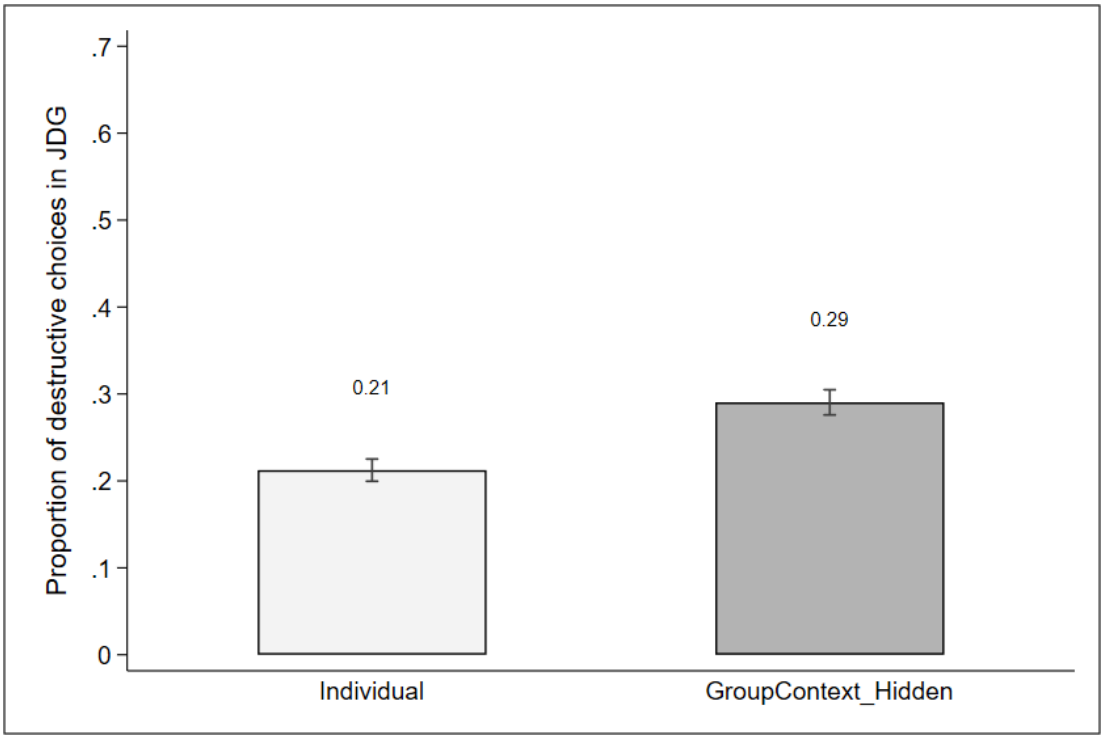
Notes: The proportion of unconditional destructive choices in the Joy of Destruction game. Bars indicate 95% confidence intervals.

FIGURE 4. DESTRUCTIVE BEHAVIOR IN GROUPS, EXPERIMENT 3 (SLOVAKIA, UNIVERSITY STUDENTS)



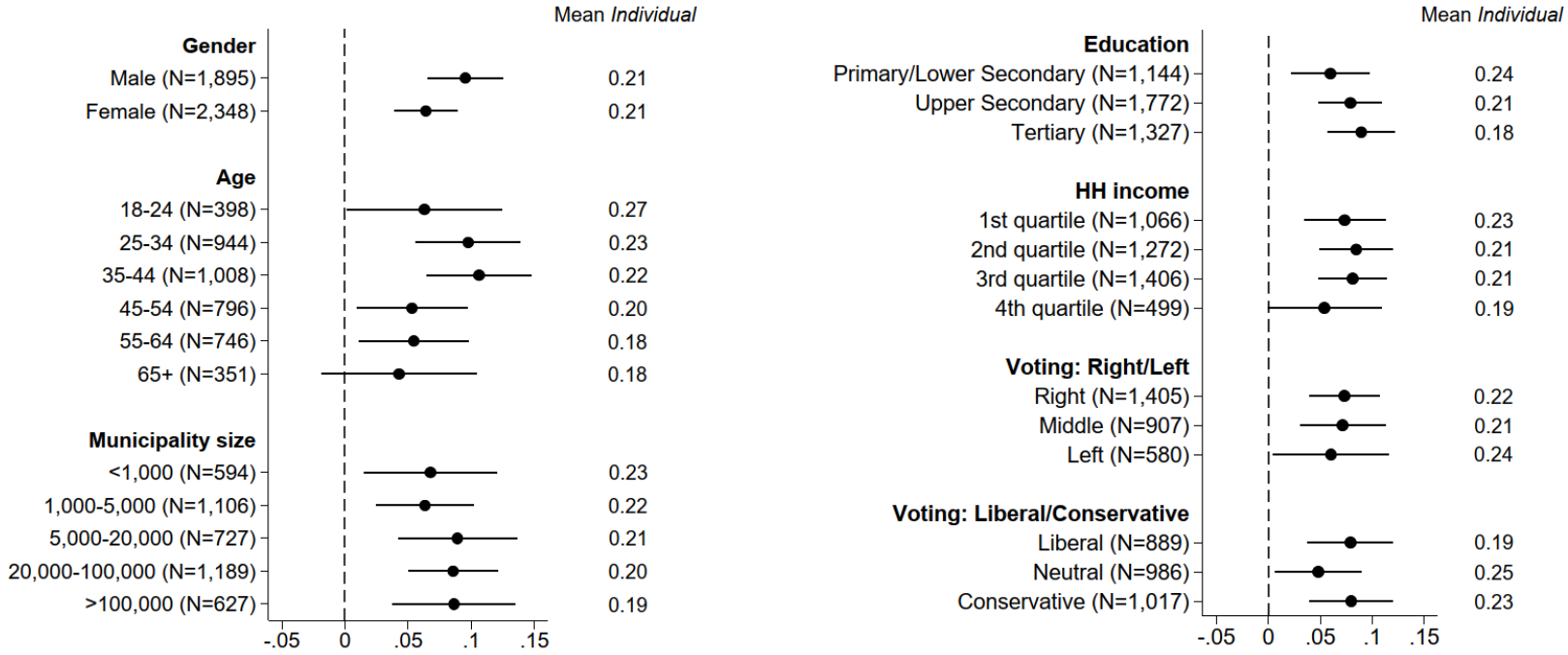
Notes: Mean index of destructive behavior in the Joy of Destruction game across conditions. The index is calculated as an average of three indicator variables for whether a decision-maker acted destructively in the unconditional choice and in two conditional choices and thus takes value between zero and one. Bars indicate 95% confidence intervals.

FIGURE 5. DESTRUCTIVE BEHAVIOR IN GROUPS, EXPERIMENT 4 (SLOVAKIA, REPRESENTATIVE SAMPLE OF ADULT POPULATION)



Notes: Mean index of destructive behavior in the Joy of Destruction game across conditions. The index is calculated 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. Bars indicate 95% confidence intervals.

FIGURE 6. EFFECT OF A GROUP CONTEXT ON DESTRUCTIVE BEHAVIOR IN THE JOY OF DESTRUCTION GAME --- HETEROGENEITY ACROSS SUB-GROUPS, EXPERIMENT 4 (SLOVAKIA, REPRESENTATIVE SAMPLE OF ADULT POPULATION)



Notes: Estimated effect of being in the *GroupContext_Hidden* condition as compared to the *Individual* condition on destructive behavior, across different demographic sub-groups. Horizontal bars indicate 95% confidence intervals. The dependent variable is an index of destructive behavior in the Joy of Destruction game, which is calculated 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. The corresponding regression table is provided in the Online Appendix (Table A.14).