Shifting Punishment on Minorities: Experimental Evidence of Scapegoating

Michal Bauer
Jana Cahlíková
Julie Chytilová
Gérard Roland
Tomáš Želinský*

Abstract

This paper provides experimental evidence showing that members of a majority group systematically shift punishment on innocent members of an ethnic minority. We develop a new incentivized task, the Punishing the Scapegoat Game, to measure how injustice affecting a member of one's own group shapes punishment of an unrelated party ("a scapegoat"). We study interactions between the majority group and the Roma minority in Slovakia. By experimentally manipulating the ethnic identity of the scapegoats, we show that the punishment "passed" on innocent individuals more than doubles when they are from the minority, as compared to when they are from the dominant group. These results illuminate a mechanism how individualized tensions can be transformed into a group conflict, dragging minorities into conflicts that are completely unrelated to their behavior.

Key words: punishment, minority groups, inter-group conflict, discrimination, scapegoating, lab-in-field experiments

JEL Codes: C93, D74, D91, J15

* 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, Czech Republic and Institute of Economic Studies, Faculty of Social Sciences, Charles University, phone: 00420224005130, fax: 00420224005333, e-mail: bauer@cerge-ei.cz; Cahlíková: Max Planck Institute for Tax Law and Public Finance, 80539 Munich, Germany; Chytilová: Institute of Economic Studies, Faculty of Social Sciences, Charles University and CERGE-EI, 110 00 Prague, Czech Republic; Roland: UC Berkeley, CEPR and NBER, 94720 Berkeley CA, USA; Želinský: Technical University of Košice, Němcovej 32, 040 01 Košice, Slovakia.

1. Introduction

How can localized social ills spiral into widespread aggressive behavior, often with tragic consequences for whole societies? Although it is widely acknowledged that aggressive behavior against ethnic minorities often arise unexpectedly and spread quickly even in previously peaceful communities (Fearon and Laitin 2000; Bardhan 2005; Esteban and Ray 2008), little is known empirically about the mechanisms that underlie these changes in behavior. In this paper we study people's biases in punishment behavior, in response to wrongdoing targeting members of one's own group. A novel aspect is that we focus on situations when punishers cannot punish the wrongdoer but only "by-standers", as in many real-life situations. Inspired by work in social psychology and political science (Doob et al. 1939; Allport 1954; Girard 1979; Staub 1992; Glick 2005), we take seriously the idea that people may have a tendency to pass revenge on innocent individuals from negatively-stereotyped minority groups. We refer to such behavior as minority scapegoating.[1,2] Such shifting of punishment on minorities violates fundamental fairness principles (Kant 1965 and classical philosophers) embedded in the legal codes of most modern societies, i.e. (i) that people should be punished only for wrongs they are responsible for and that they intentionally committed, and (ii) that all people should receive equal punishment for the same wrongdoing. Demonstrating such bias in

[1] The term "scapegoat" has biblical origins, coming from the Book of Leviticus—it refers to a sacrificial goat to which the ancient Israelites symbolically transferred their sins. A scapegoat was originally understood as a person or an animal that symbolically absorbed sins of others and carried them away from those who committed them (Crossman 2018). According to social psychology, scapegoating occurs when punishment of the true source of the anger is inhibited and people shift their aggression towards other individuals (Allport 1954; Glick 2005; Newman and Caldwell 2005). Staub (1992, p. 48) describes this behavior as follows: "*When there is no aggressor or the aggressor is too powerful or the source of responsibility cannot be identified or the responsibility is one's own (or one's group), identifying a scapegoat will have "beneficial" psychological effects*" Scholars have also suggested that such displaced punishment typically targets individuals or groups who are weak and vulnerable, as well as already deprecated minority groups, perhaps because it is easier to *ex post* (irrationally) blame them for one's own misfortunes and to attribute negative imputed characteristics to justify aggression against them (Bettelheim and Janowitz 1950).

[2] Minority groups are targets of violence in many parts of the world (Horowitz 1985; Yanagizawa-Drott 2014). Historical evidence suggests that aggressive attitudes and behaviors, including pogroms and attempted genocides, increase during periods of social unrest and economic problems within the majority group (Anderson, Johnson, and Koyama 2017; Grosfeld, Sakalli, and Zhuravskaya 2019; Voigtlander and Voth 2012). Survey evidence from the US and Europe indicates that hostility towards 'other' groups, especially weaker groups like migrants or homosexuals, becomes activated when social cohesion within the majority group starts to fall apart, creating a sense of normative threat (Stenner 2005; Stenner and Haidt 2018).

punishment preferences of a dominant groups is important, because it may drag minorities into conflicts that are completely unrelated to their behavior and transform individualized tensions into a group conflict.

To study this behavioral phenomenon, we develop a novel experimental paradigm, the *Punishing the Scapegoat Game*. In this game, impartial spectators from the majority ethnic group can impose a monetary punishment on others at own cost, after observing that someone malevolently destroyed the earned income of an individual from their own group. Importantly, the punishers cannot punish the wrongdoer, but they can punish a "bystander" (a scapegoat) who does not know the wrongdoer and who is not involved in any way in the original wrongdoing. By manipulating the identity of the passive scapegoats, we show that a desire to "pass" the punishment on innocent individuals more than doubles when such persons are from an ethnic minority, as compared to when by-standers are from the dominant group. In addition, we document that when spectators can punish actual wrongdoers, they punish wrongdoers from the minority more harshly than members of their own group, in line with earlier work (Bernhard, Fischbacher, and Fehr 2006; Schiller, Baumgartner, and Knoch 2014). Together, we establish that biases in punishment preferences against the minority group we study are not limited to situations when members of the minority are responsible for harmful actions, but, importantly, mark behavior also in situations when it is clear that minority members have nothing to do with the social ill.

Identifying people's desire to pass punishment on innocent individuals, and its determinants, is empirically challenging with naturally-occurring data. First, it is nearly impossible to rule out the role of the standard economic incentives to harm innocent individuals, such as simple self-interested grabbing of resources and economic power from them. In addition, in most real-life situations there is an element of uncertainty about who originated the harm. Members of the dominant group may punish innocent individuals from minority groups because they over-attribute responsibility for misfortunes to actions of minority groups. Controlled experimental environment allows closing these confounding mechanisms, and focusing on whether shifts in punishment onto minorities are embedded in preferences or decision heuristics. Specifically, our aim is to identify indirect *negative* reciprocity -- an individual desire to engage

in hostile behavior towards unrelated parties, in response to experienced hostility against self or someone one cares about. In light of the literature on scapegoating, the key question is whether indirect negative reciprocity becomes stronger when the unrelated party (by-stander) is a member of a minority group. Answering this question requires an empirical setup that allows a researcher to (i) measure how people behave when punishment of wrongdoers is inhibited and people can punish only individuals who could not have causally contributed to the original harmful act, (ii) measure punishment responses in one-shot anonymous interactions that are costly for the punisher, and that provide no scope for material benefits of punishment, and (iii) compare behavior towards a weaker minority group and towards the own group.

The *Punishing the Scapegoat Game* aims to address these challenges. The game builds on the Third Party Punishment Game[3], and elicits behavioral responses to observations of unfair behavior against a member of one's own group. While existing incentivized experiments on punishment of socially undesirable behavior focus exclusively on direct punishment of individuals who make active decisions whether or not to violate a social norm, a key element of the Punishing the Scapegoat Game is that we add a fourth person, the passive Scapegoat. This feature allows us to separate the person who commits a harmful act and a person whom the impartial spectator can punish. We exogenously manipulate information about the ethnicity of the Scapegoat to test whether people are more prone to redirect the punishment onto minority group members, if they cannot punish the Wrongdoer.

We study the interaction between real ethnic majority and minority group members and focus on young adults, because earlier research indicates that perpetrators of aggressive behavior against other ethnic groups, including hate crimes, are typically young (e.g., Levin and McDevitt 2002). The setting of our study is Eastern Slovakia, a region experiencing high unemployment, where the majority population lives side-by-side with an economically disadvantaged, segregated, and negatively stereotyped minority (Roma), who represent approximately 15% of the local population. Scattered across many European countries, including

[3] The Third Party Punishment game has been originally developed by (Fehr and Fischbacher 2004) and later became one of the prominent experimental tasks to study direct punishment of norm-violators, see, for example, (Bernhard, Fischbacher, and Fehr 2006; Goette, Huffman, and Meier 2006; Henrich et al. 2010).

Slovakia, the Czech Republic, France, Italy, and Hungary, Roma people are the largest ethnic minority in Europe. Thus, Eastern Slovakia is an apt natural setting to test whether aggression by the majority group can be easily displaced towards members of a minority group that is relatively weak and negatively stereotyped.

Before the main experiment, we organized a supplementary activity in which a sample of workers received equal earnings for performing a tedious work assignment. A randomly selected subgroup of workers, the Wrongdoers, then made an anonymous decision whether to reduce the earnings of another group of workers, the Victims, with no pecuniary benefits to themselves. This work stage creates a real sense of entitlement on the part of the workers to the earnings provided, making the decision to destroy the earnings of the Victims very anti-social. This destructive behavior is presented to experimental subjects (Punishers), along with the information that the Victim is from their own ethnic group. We focus on the punishing behavior of spectators (Punishers), rather than the direct Victims. Indeed, if we want to understand whether scapegoating behavior can contribute to the spreading of aggressive behavior beyond parties directly involved in the original wrongdoing, it is more relevant to identify a scapegoating response among individuals tied to the victim *only* by shared group membership.

The main outcomes of interest are the decisions of the spectators (Punishers) from the majority ethnic group. We measure how they respond to information on whether a Wrongdoer caused harm to a Victim and how serious the harm was. Punishers decide whether and to what extent to reduce the earnings of a Scapegoat who is completely passive, i.e. neither causes the harm nor is in a position to prevent it, and who does not know the Wrongdoer. The punishment decision is one shot, anonymous, and costly and the Wrongdoer is not informed about the punishment of the Scapegoat. Therefore, by design, punishing a Scapegoat cannot be explained by self-regarding motives, strategic considerations, such as deterrence of future offences, or the provision of moral education to the Wrongdoer. Further, in an additional condition the Punishers can punish the Wrongdoer.

The main results are the following. First, regarding the direct punishment of the individual (Wrongdoer) who makes an active decision whether to harm a Victim or not, we find that members of the minority group receive harsher punishment than members of the majority group. The same unfair behavior triggers a magnified revenge against members of the Roma minority as compared to the Slovak majority group.

Second, when direct punishment of Wrongdoers is not possible, we show that a non-negligible fraction of Punishers (23%) redirect punishment onto Scapegoats. Importantly, identity of the Scapegoat matters: the punishment is twice as severe when the Scapegoat is from the Roma minority than when the Scapegoat is from the majority group.

Third, we show that the magnified tendency to punish the minority Scapegoat is not driven by collective punishment, i.e. by the case in which both the Wrongdoer and the Scapegoat are from the minority group. We find that the minority Scapegoats are punished more often and more severely than majority Scapegoats even when the original harm is committed by a member of the majority group, and thus when Wrongdoers and Scapegoats do not share the same ethnicity.

Next, we document that discrimination against the Roma minority is *activated* by the unfair behavior of the Wrongdoers. When Wrongdoers do not harm a Victim, ethnic majority Punishers do not discriminate against the ethnic minority. However, when Punishers observe an injustice against a member of their group, the Roma minority subject pays a disproportionate price, and discriminatory punishment arises. Therefore, this set of results indicates that the observed biases in punishment do not originate in unconditional hostility against the minority that would manifest under any circumstances. Rather, they support the interpretation that they originate in magnified preference to retaliate (negative reciprocity) when interacting with the minority, both when punishing the wrongdoers as well as the bystanders (indirect reciprocity). Put differently, discriminatory preferences are latent in "peaceful" times, but are activated in environments when decision-makers respond to social ills, and the majority group may shift punishment for such social ills onto minorities.

Finally, we study whether shifting of punishment on another group characterizes decisions of the dominant group towards a weaker minority, or also vice versa. To do so, we measure the behavior of Roma Punishers in the same experimental set up and compare their punishment behavior with Punishers from the majority group. We find weaker evidence of biases in punishment against the majority group, as compared to the biases of the majority group against the minority group. This is in line with prior evidence, suggesting that shifting of punishment onto innocent individuals is psychologically easier when the target is a member of a negatively stereotyped, vulnerable, and smaller group (Bettelheim and Janowitz 1950; Allport 1954).

Our work contributes to several literatures. A long-standing literature has studied the existence of ethnic discrimination and in-group favoritism in normal circumstances.[4] This paper is related to a smaller body of work that aims to identify situational factors that can trigger or magnify discrimination and hostility against ethnic minorities. Earlier work has focused on the institutional environment (M. Alexander and Christia 2011), proximity to terrorism (Shayo and Zussman 2011) and to violent elections (Hjort 2014; Berge et al. 2020). In the same setting that we study here, it has also been shown that conformism within salient social groups of adolescents can give rise to hostility against the Roma minority (Bauer et al. 2018). This paper focuses on biases in people's tendency to pass revenge onto innocent individuals, triggering discriminatory behavior in situations when they cannot punish the true source of injustice. This relates our work to the literature on indirect reciprocity. Indirect positive reciprocity (rewarding kind acts with kind acts towards other individuals) has been described as a mechanism that promotes spreading of pro-social behavior (R. Alexander 1987; Nowak and Sigmund 2005). The existing experiments document existence of indirect positive reciprocity not only in an environment with repeated interactions, where it can be motivated by strategic selfish considerations, but also when there is no scope for reputation-building

[4] One stream of the empirical literature has measured existence of discrimination based on real-life group attributes, such as ethnicity (e.g., (Fershtman and Gneezy 2001; Bertrand and Mullainathan 2004)), similarly as we do in this paper. The existence of in-group favoritism and out-group discrimination have also been measured using a minimal-group experimental paradigm, by creating artificial group boundaries in the laboratory (based on, for example, having T-shirts of the same color or sharing preferences for art) and the most prominent examples are Tajfel (1981) and Chen & Li (2009).

(Bolton, Katok, and Ockenfels 2005; Engelmann and Fischbacher 2009), including one-shot interactions (Stanca 2009; Dufwenberg et al. 2001). We contribute to this literature by focusing on indirect negative reciprocity (responding to hostile acts, by engaging in hostile behavior towards unrelated parties), in order to understand better the conditions under which (inter-group) conflicts spread. We also show that the identity of the unrelated party plays an important role.

Our experimental design is most closely related to existing lab-in-field experiments on punishment of norm violations across group boundaries (Goette, Huffman, and Meier 2006; Bernhard, Fischbacher, and Fehr 2006). Earlier work has made progress in studying group biases in the direct punishment of active norm violators. In particular, among indigenous tribes in Papua New Guinea, Bernhard et al. (2006) document that third parties who share group membership with victims punish out-group members more than in-group members, which is similar to the patterns of Wrongdoer punishment that we find.[5] We contribute by developing an experimental task that sheds light on determinants of the punishment of passive individuals, to test whether revenge may spill-over and affect behavior not only towards a wrongdoer but also unrelated individuals.[6] Interestingly, both (Bernhard, Fischbacher, and Fehr 2006) and (Goette, Huffman, and Meier 2006) find that people punish more severally wrongdoing targeting members of their own group, as compared to members of an out-group. This result motivated us to keep the identity of the victim fixed, and to focus on punishment responses for wrongdoing against own ethnic group.

Further, our results speak to the ongoing debate about whether giving people opportunities to punish leads to desirable social outcomes or not. Many studies have documented that punishment can have a pro-social norm enforcement function (Fehr and Gächter 2002; Gächter, Renner, and Sefton 2008;

[5] Focusing on a different type of real-life social groups (based on poticial party affliation, soccer club support and membership to different platoons in Swiss Army), similar pattern has been found in Schiller, Baumgartner, and Knoch (2014) and Baumgartner et al. (2012), but not in Goette, Huffman, and Meier (2006) and Goette et al. (2012).

[6] Another interesting line of experiments with a punishment option studies the attribution of responsibility and shifts in blame caused by the diffusion of responsibility for unkind actions across multiple individuals (Bartling and Fischbacher 2012). Here, we study redirecting punishment towards individuals who are completely passive. Responsibility for unkind behavior can thus not be attributed to them.

Henrich et al. 2010). Recent work has started to explore the conditions in which providing an opportunity to punish may not have positive social effects (Herrmann, Thoni, and Gächter 2008; Abbink et al. 2010; Goette et al. 2012). For instance, Herrmann et al. (2008) find that, in some societies, a non-negligible fraction of people punish individuals who are relatively cooperative (a behavior called 'anti-social punishment'), and that such behavior eliminates the pro-social effects of punishment. We show that anti-social punishment and scapegoating are two distinct behavioral phenomena: while we find no group bias in anti-social punishment (i.e., people do not punish cooperators from the other group more than their ingroup members), we find systematic biases against the minority in the punishment of the Scapegoat.

2. Experimental Design

The experimental design is presented in six subsections. First, we provide a short background on the Roma ethnic minority. Second, we describe the sample selection. Third, we present the tasks used to measure punishment of harmful behavior performed by a third party—either by directly punishing the Wrongdoer or by punishing the Scapegoat, who is an uninvolved bystander. Fourth, we describe how we manipulate information about the identity of the Wrongdoer and the Scapegoat. Fifth, we describe how we elicited real wrongdoing from a supplementary sample of individuals who were given the option to malevolently reduce the earned income of other individuals. Last, we provide further details about our experimental procedures.

2.1 Background on the Roma ethnic minority

Eastern Slovakia represents an apt natural setting to study behavior of the dominant group towards a negatively stereotyped and weaker ethnic minority. The Roma people, a minority of Indian origin, constitute the largest ethnic minority in Europe, estimated at 10-12 million persons.[7] They live in generally poor socio-

[7] The background section draws heavily on our earlier paper (Bauer et al. 2018) about the Roma, which uses a different data set collected in the same setting.

economic conditions and experience social exclusion all over Europe.[8] The average education levels of Roma are low (20% finish upper-secondary education), they are poorly integrated into labor markets (less than one third are in paid employment), they generally live in substandard housing, and they have worse health and lower life expectancy than the majority populations. It is estimated that 85% of Roma in Europe live below national poverty lines. In Eastern Slovakia, the setting we study, the Roma represent around 15% of the local population. Around 65% of Slovakian Roma live segregated from the majority population, often in isolated settlements or on the edges of villages and towns. Previous research shows that Roma are subject to prejudice and face discrimination in labor and housing markets (Bartoš et al. 2016). According to reports by the European Commission, almost one quarter of Europeans (38% of Slovaks) state that they would be uncomfortable having a Roma neighbor and 34% of Europeans (60% of Slovaks) think citizens in their country would feel uncomfortable about their children having Roma classmates.

Since World War II, when the Roma were persecuted along with the Jews, there has not been any systematic violent conflict involving the Roma. Nevertheless, the frequency of anti-Roma violence has been increasing in the last decade, especially in Central and Eastern Europe (Council of Europe 2012). During the last decade, a series of anti-Roma marches have been staged in dozens of towns and cities across the region. These protests, in which far-right extremists have often been joined by members of the local population, have commonly escalated to property damage and/or violence. In addition, walls separating the majority population from their Roma neighbors have been built in numerous cities in Slovakia since 2008, and also in other countries including Romania and Bulgaria.

[8] The most significant populations are in Central and Eastern European countries (Bulgaria, Romania, Slovakia, Hungary and the Czech Republic), but Roma also live in France, Greece and Italy.

2.2 Sample selection

The data collection took place between May and September 2017, across various localities in Eastern Slovakia. We study punishment behavior among a diverse (though not representative) sample of young adults of the majority Slovak ethnicity (N=337).

The subjects are students aged 18-23 from the last two grades of secondary school (57% of the sample) and a local university (43%). The secondary school students come from seven different schools, located in two regions—Košice and Prešov—and include students from the three most common types of high schools in Slovakia: general, technical, and business/commerce. In each region, we randomly selected at least one school of each type; all schools selected agreed to participate. The school headmasters determined which classes would take part. We randomly selected participants from all interested adult students (aged 18 and above) in those classes and the experiments were implemented in their respective schools. University students were recruited at the campus of the Technical University of Košice, and therefore the vast majority were enrolled in technical or economic majors. The sample characteristics are presented in Table A1 (in Supplementary Online Appendix A): the subjects are on average 19.3 years old, and 42% are female. Their parents completed at least secondary education in almost all cases; 25% of fathers and 32% of mothers also hold a university degree.

Further, we study the behavior of a sample of young adults from the Roma minority (N=484)[9]. We randomly selected 21 villages and towns with an estimated Roma population of over 1,000 (the list of eligible communities was based on the *Atlas of Roma Communities 2013*), across the same two regions. We contacted a village representative, typically a mayor, asking for permission to conduct research and to appoint a local contact person who was responsible for inviting participants (of 42 villages contacted, 21

[9] We have intentionally selected a larger sample than from the majority group, because piloting of the experiment revealed that around 25% of the Roma sample did not properly understand the experimental instructions, based on crosscheck questions on understanding. In the analysis, we report results for the whole sample, as well as for a sub-sample that excludes individuals whose responses indicated limited understanding of the instructions. The results are robust.

agreed to take part). The local coordinators were instructed to select approximately 25 participants per village, Slovak-speaking, 18-24 years of age, literate, ideally living in different parts of the village and belonging to different social groups within the community. The experiments took place in local community centers[10]. As expected, the Roma subjects had less education and came from a poorer socio-economic background (Table A2) compared to the subjects from the majority group: 15% were students, 49% were unemployed and among those no longer studying, 55% had finished only primary school. Parents of the Roma subjects had typically completed primary school at most.

2.3 Experimental tasks

Each spectator (Punisher) makes decisions in two tasks: (i) the *Punishing the Wrongdoer Game*, which is similar to the standard Third-Party Punishment Game, and (ii) in a newly designed *Punishing the Scapegoat Game*. The tasks are illustrated in Figure 1. In both tasks, a Punisher is matched with three people – a Wrongdoer, a Victim and a Scapegoat, who are neutrally labeled Person A, Person B and Person C when described to the decision-maker. Specifically, on a tablet computer, the Punisher observes three pictures: the first one with twenty potential Wrongdoers, the second one with twenty potential Victims, and the third one with twenty potential Scapegoats. Each picture displays twenty passport-style photos, homogenous in terms of ethnicity, and taken against a neutral background. The Punisher knows that s/he is matched with one person from each set of twenty photographs but does not know with whom specifically. The Punisher is informed that each of these three people (Wrongdoer, Victim, and Scapegoat) completed a work assignment and earned 8 euro for their work. Further, the Punisher learns that after completing their work, the Wrongdoer had an option to reduce the earnings of the Victim by 0, 2, 4, 6 or 8 euro, and that the Scapegoat was utterly passive. Punishers learn that the Wrongdoer, the Victim, and the Scapegoat are each

[10] In one village, the experiments took place in several houses within the village, since the community center was unavailable.

from a different location, do not know each other, and that only the Wrongdoer had the option to reduce the earnings of someone else, and only the Victim's money could have been reduced.

[Figure 1 around here]

In the *Punishing the Wrongdoer Game*, the Punisher can reduce the payoff of the Wrongdoer by 0, 2, 4, 6, or 8 euro. In the *Punishing the Scapegoat Game*, the Punisher cannot punish the Wrongdoer but can reduce the payoff of the Scapegoat. Thus, the *Punishing the Scapegoat Game* mimics the definition of scapegoating: Punishers face a situation when their ability to address the true source of the social problem faced by own group is inhibited – they cannot punish the Wrongdoer -, but they can instead displace their anger onto a different person and treat him as a Scapegoat.

In both tasks, punishment is costly: reduction of each euro costs the Punisher 0.10 euro. Punishers' decisions are elicited for all five possible actions of the Wrongdoer towards the Victim, using a strategy method. When making the decision, in a simple tablet interface, the Punishers choose a preferred punishment level for each possible action of the Wrongdoer by pushing a "minus" or "plus" button—each tap on the tablet leads to a 2.00 euro decrease/increase of the earnings of the Wrongdoer or the Scapegoat. After each tap, the payoffs of the Wrongdoer, the Victim, the Scapegoat, and Punisher's own payoff are updated and clearly specified on the screen (sample decision screens are depicted in Figure A1). The decision about the Wrongdoer's payoff was framed as a punishment, whereas the decision about the Scapegoat's was framed neutrally. In order to avoid any scope for instrumental punishment of the Scapegoat, the Punisher knew that the Wrongdoer would not be informed of the decision affecting the Scapegoat. The full experimental protocol is provided in the Supplementary Online Appendix B.

The two tasks described above were conducted in random order, and the participants did not know about the existence of the second task until after they finished the first one. Each task was payoff-relevant

at 10% probability (but the tasks could never both be payoff-relevant)[11], in which case the real decision of the Wrongdoer together with the Punisher's choice in that scenario determined the final payoffs of the Wrongdoer, the Victim, the Scapegoat and the Punisher. In order to avoid aversion to punishment motivated by disadvantageous inequality, we set the Punisher's endowment at 9 euro and thus even if s/he chose the maximum punishment level, his/her final payoff was 8.20 euro, i.e. higher than the payoff of all other players.

2.4 Manipulating the identity of Wrongdoer and Scapegoat

The Victim is always of the same ethnicity as the Punisher. Importantly, we exogenously manipulate signals of ethnicity of the Wrongdoer and of the Scapegoat, in order to identify how ethnicity affects the decisions of the Punishers, as illustrated in Figure 1. In the Wrongdoer SAME condition, the Wrongdoer has the same ethnicity as the Punisher, whereas in the Wrongdoer OTHER condition, the Wrongdoer comes from the other ethnic group. Similarly, in the Scapegoat SAME condition, the Scapegoat is of the same ethnicity as the Punisher, while in the Scapegoat OTHER condition, the Scapegoat comes from the other ethnic group. The signals of ethnicity of the Wrongdoer and of the Scapegoat are manipulated orthogonally.

In this 2x2 "between-subject" design, each Punisher is randomly allocated to one of the four possible combinations of Wrongdoer SAME/OTHER and the Scapegoat SAME/OTHER conditions. Thus, participants did not make allocations towards two Wrongdoers (and Scapegoats, resp.), once in SAME and once in OTHER. This implies that we cannot estimate discrimination at the individual level, but an important advantage of this design choice is that it blurs the purpose of the study (measuring discrimination)

[11] To make both tasks incentive-compatible in this way, at the end of the experiment, one ball was picked at random from a bag containing ten balls. Before starting Task 1, the Punishers knew that there was a star on one of the balls and Task 1 would be payoff-relevant if this ball was selected. After learning about the existence of Task 2, subjects were told that there was a triangle on another of the ten balls, and if that ball was selected, Task 2 would be payoff-relevant.

and mitigates concerns about experimenter demand effects affecting the estimated discrimination. Randomization checks indicate that the randomization was successful (Column 7 of Tables A1 and A2).

We use photographs to signal ethnicity. Photographs provide a clear signal of a group attribute, because Roma people (who are of Indian origins and have a darker skin color) are visually distinct from the Slovak majority (white). For each group of workers, we displayed pictures of twenty individuals instead of one, in order to avoid a case in which sympathies/antipathies towards a specific person in a picture might drive decisions. Also, the individuals displayed in the pictures were homogenous in terms of gender and age (all were young male, 18-24 years old), in order to avoid differential treatment based on these attributes.

Since we study differential behavior towards real groups, note that the pictures in SAME and OTHER may signal characteristics other than ethnicity, particularly socioeconomic status. Thus, a natural question is whether our design measures discrimination against other ethnic groups, or towards poorer or less educated individuals. To partially address this issue, we asked each Punisher, after making their experimental choices, about their perceptions of socioeconomic status (employment status, father's education) of individuals displayed as Wrongdoers and Scapegoats. We show that our results are robust to controlling for differences in perceptions across the SAME and OTHER conditions.

2.5 Elicitation of wrongdoing

Prior to the punishment experiment (May and early June 2017), we organized the following supplementary work activity among a sample of different individuals, in order to make the situation a Punisher was presented with real, including the harm committed by the Wrongdoer to the Victim, and to make the Punisher's choices consequential.

We recruited workers from the majority group and workers from the Roma ethnic group. Since the experimental design requires six distinct groups of individuals who do not know each other (majority Wrongdoers, Roma Wrongdoers, majority Victims, Roma Victims, majority Scapegoats and Roma

Scapegoats), we recruited workers in six different localities. In each locality, we hired 20-23 workers, all young men, aged 18-23, and a professional photographer took a passport photo of each worker.[12] We selected the photographs of twenty workers and composed a picture consisting of these twenty photos. The pictures were used to signal the ethnicity of the Victim to the Wrongdoers and in the main experiment to signal the ethnicity of the Wrongdoer, the Victim and the Scapegoat to the Punishers.

Workers performed eight hours of a simple and useful task, aimed at improving the local environment (e.g., cleaning; painting desks; painting a fence; collecting and sorting tree boughs in the woods) and could earn up to 8 euros per hour of work. In the experiment, we treat each hour of work as a separate work assignment. Thus, eight different work hours of the same worker enter decisions of eight different Punishers. This procedure ensures that each Punisher observes a real situation in which the Wrongdoer, the Victim and the Scapegoat displayed in the pictures worked for one hour and could earn up to 8 euros for this work assignment. The Punishers were not informed that the workers worked for more than one hour.[13]

Upon the completion of the work assignment, each Wrongdoer made a decision whether to reduce up to 8 euro of the potential earnings of the Victim. Specifically, they could leave Victim's earnings for an hour of work as they were, or lower them by 2, 4, 6, or 8 euro. Since we wanted to create a strong urge to punish among Punishers, we deliberately elicited a particularly malevolent form of harmful behavior: it reduces the earned income of the Victim and does not create a pecuniary benefit for anyone, including the

[12] The workers signed a consent form, agreeing to their participation in research and the use of their photograph for the purposes of the research.

[13] Workers from two localities (one majority and one Roma) were assigned to the position of Victims for all eight work hours. Pictures from these two localities were always used to display a potential Victim. Workers from the remaining four localities were assigned to the position of Wrongdoers for four work hours (and made four decisions whether to reduce earning of a Victim) and to the position of Scapegoats for the remaining four work hours. Pictures from these localities were randomly assigned to display either the potential Wrongdoer or the potential Scapegoat, in order to ensure that specific photographs in each of the pictures do not drive the differences between the punishment of the Wrongdoer and the Scapegoat.

Wrongdoer.[14] We find that a non-negligible fraction of Wrongdoers chose to do harm (see the histogram in Figure A2), in line with earlier work documenting a relatively large prevalence of anti-social behavior (Abbink and Sadrieh 2009; Prediger, Vollan, and Herrman 2014; Bauer et al. 2018) and also in line with beliefs about prevalence of harming, elicited among Punishers, as described below.

The Wrongdoers knew that their decisions would have real payoff consequences, and would be anonymous: the Victim would be given the remaining amount without any additional information. They also knew that the Victim could not reduce their (the Wrongdoer's) wage. Further, they were informed (as all other workers) that their own earnings could be affected by decisions of other people but did not know any details about the Punisher's task.

Given the expected destruction rates among the Wrongdoers and punishment rates among the Punishers, we could guarantee to each worker the payment of 20 euros shortly after they finished the work. This amount is equivalent to the minimum wage in Slovakia at the time of the data collection (2.50 euro per hour). The workers were further informed that they would receive up to 44 additional euro in three months based on the decisions of other people. Thus, the maximum payment they could receive was 64 euro, equivalent to 8 euro per hour of work. The average amount paid to workers was 45.3 euros and the minimum amount was 40 euros.

2.6 Procedures

We paid particular attention to maximize correct understanding of the tasks. First, the experimenters explained the instructions one-on-one. Second, the working/wrongdoing stage of the experiment and the

[14] Note that in the standard Third Party Punishment game, the players receive resources as "manna from heaven", while in our experiment, inspired by recent experiments on redistribution of earned income (Cappelen et al. 2007; Almas, Cappelen, and Tungodden 2019), we add the work stage, which creates a sense of entitlement, after which the Wrongdoer destroyed earned resources of the Victim. Also, while prior experiments study punishment of violators of norms that govern the positive side of social behavior (Goette et al., 2006, 2012), we focus on the responses of Punishers to unambiguously nasty behavior—destruction of the Victim's earnings.

punishment choices were explained in detail using a simple tablet interface with the photos of Wrongdoers, Victims, and Scapegoats (Figure A1).[15] Third, before making decisions, the Punisher had to answer six comprehension questions, demonstrating understanding of the roles of the three types of workers. If any of the answers were not correct, the experimenter explained the whole setup once again, and the comprehension questions were asked a second time. As a robustness check, we exclude from the sample individuals who did not provide correct answers when the questions were asked for the second time.

The Punishers made decisions in the *Punishing the Wrongdoer Game* and in the *Punishing the Scapegoat Game,* in randomly chosen order. The tasks were explained in detail and subjects went through examples to familiarize themselves with the tablet environment. To avoid perception of social pressure, the experimenters could not observe subjects' decisions. For each possible action of the Wrongdoer, the experimenter described the situation, and then gave the subject privacy to anonymously make decisions on the tablet computer whether and how much to lower the payoffs of the Wrongdoer (in the Punishing the Wrongdoer Task) or the Scapegoat (in the Punishing the Scapegoat task).[16]

After completing the main tasks, we elicited beliefs about the action of the Wrongdoer, asking the Punishers to guess which of the five options the Wrongdoer chose (subjects were rewarded 1 euro for a correct guess). For the sample of majority Punishers, we also elicited beliefs in more detail, asking how many of the twenty potential Wrongdoers they thought chose each action, to gauge beliefs about the distribution of responses. We did not elicit the more detailed beliefs for the Roma minority Punishers, because the explanation was too time consuming and a large percentage of subjects did not understand this task properly during the pilot.

A short post-experiment questionnaire followed. To elicit the perceptions of the Wrongdoer's and Scapegoat's social status, the Punishers were asked to guess their employment status

[15] The app was programmed using z-Tree (Fischbacher 2007).

(student/employed/unemployed) and the educational level of their father. Subjects also filled in a short questionnaire on their own demographics (gender, education, family size, marital status, parental education and employment, religiosity, and household amenities).

To determine final payoffs, at the end of the experiment, each Punisher chose one ball out of ten at random from a bag, which determined whether their decision in the *Punishing the Wrongdoer Game* and in the *Punishing the Scapegoat Game* was payoff relevant or not. An actual decision of an anonymously matched Wrongdoer determined which decision situation in a given task was payoff relevant (out of five). Punishers received their earnings in cash immediately. Workers received a part of their earnings after completing their work and the remainder a few days after completing the punishment experiment.

The experiments were always conducted in private; other persons could not hear the instructions or see the decisions that the subjects made on the tablets. The experiment typically lasted 25 minutes per subject. On a typical day, three experimenters implemented the experiments in one location. Each subject was randomly allocated to one of the experimenters and we control for experimenter fixed effects (ten experimenters overall) in the estimates.[17]

3. Results

In the main analysis, we focus on decision-makers from the majority group and how they behave towards the minority group. In Section 3.1, we analyze punishment of the Wrongdoer, i.e. direct punishment of the person who made an active decision to harm the Victim by reducing her earnings. In Section 3.2, we study the punishment of the Scapegoat, i.e. a person who did not do any harm to the Victim. In Section 3.3, we provide additional results, including the analysis of decision-makers from the Roma minority, and discuss

[17] All experimenters were from the majority ethnic group, and they were randomly allocated to conditions within each location. The results are robust to controlling for experimenter fixed effects. Thus, differences in characteristics of experimenters cannot explain the discriminatory patterns.

how our findings relate to minority scapegoating and other existing theories of discrimination (collective responsibility, statistical discrimination, and taste-based discrimination).

3.1 Punishment of the Wrongdoer

Are members of the Roma minority punished more severely than members of the majority group? Figure 2 displays the results for direct punishment of the Wrongdoer across the five specific amounts of Victim's earnings that the Wrongdoer could decide to destroy (0, 2, 4, 6 and 8 euros). Panel A shows the average amount of euros that the Punishers decided to destroy in the SAME condition, when the Wrongdoer is also from the majority population, and in the OTHER condition, when the Wrongdoer is a member of the Roma ethnic minority. Panel B depicts the proportion of the Punishers who decide to punish in each case. Table 1 presents the results of the regression analysis.

[Figure 2 around here]

We find several interesting patterns. First, the Punishers are sensitive to the intensity of the harm done by the Wrongdoer to the Victim. When no harm is caused, the extent of punishment is negligible (0.24 euro on average out of 8 euro maximum). The greater the harm caused by the Wrongdoer, the stronger the punishment. On average, each additional euro of a Victim's earnings destroyed by the Wrongdoer increases the punishment by 0.64 euro.

Second, in "normal" circumstances, when no harm is done by the Wrongdoer, we do not detect any discrimination against the Roma minority. If anything, the decision-makers destroy higher amounts of the earnings of subjects from their own ethnic group—the average extent of punishment is 0.13 euro in the OTHER condition vs. 0.35 euro in the SAME condition.

Third, the sensitivity to the extent of the harm is systematically larger in OTHER as compared to SAME. In SAME, when a Wrongdoer destroys one more euro of the earnings of the Punisher's ingroup member, the average extent of punishment increases by 0.56 euro, while in OTHER, it goes up by 0.71

euro. The coefficient for an interaction term between harm intensity and OTHER is statistically significant at the 1% level (p=0.002, Column 1 of Table 1). The results are robust to using various specifications and control variables, including order of the task, experimenter and location fixed effects (Table A3).

[Table 1 around here]

Greater harm intensity increases both the likelihood that subjects will punish the Wrongdoer and the extent of the punishment among those who decide to punish (Columns 2 and 3 of Table 1). When there is no harm, the proportion of Punishers who reduce Wrongdoer's earnings is around 8%,[18] and this fraction increases to around 78% when Wrongdoers decided to destroy 4, 6 or 8 euros of victim's earnings (Panel B of Figure 2). Again, the effects are stronger in OTHER than in SAME as indicated by the interaction term between harm intensity and OTHER, which is significant at the 1% level both for the extensive and intensive margins. Specifically, each additional euro of the Victim's earnings destroyed by the Wrongdoer leads to an increase in the proportion of those who decide to sacrifice part of their own payoff in order to punish the Wrongdoer by 7 percentage points in SAME and by 9 percentage points in OTHER, and among those who punish, to an increase in the extent of punishment by 0.63 euro in SAME and by 0.77 euro in OTHER.

Finally, we show that this magnified revenge in OTHER as compared to SAME gives rise to discrimination against the Roma minority (Figure 3). When the harm is large (4, 6 or 8 euros), Wrongdoers from the minority group are punished more harshly, by 13-20%, than Wrongdoers from the decision-makers own group (p-values = 0.051, 0.095, 0.013 respectively). In the situation in which the Wrongdoer committed maximum harm (destroyed all 8 euros), the punishment gap is the most profound (4.75 euros in SAME vs. 5.69 euros in OTHER, on average).

[Figure 3 around here]

[18] The proportion of Punishers who reduce Wrongdoers' payoff when no harm was caused is substantially smaller than the proportion of Wrongdoers who chose to reduce Victims' earnings in the supplementary activity. Note, however, that the two decision-making situations are not directly comparable, one important difference being the fact that for the Wrongdoer the decision to reduce other's earnings was costless while for the Punisher it came at a cost to self.

To sum up, when no harm is done to a member of the majority group, the decision-makers from the majority group do not discriminate, i.e., do not harm the ethnic minority more often. However, Wrongdoers from the Roma ethnic minority are punished more severely than Wrongdoers from the majority population, for the same harmful actions.

3.2 Punishment of the Scapegoat

In this sub-section, we explore behavior towards the Scapegoat. In this situation, the subjects cannot directly punish the Wrongdoer. Instead, they can lower the payoff of the Scapegoat—a person who is from a different village and completely uninvolved in the harm committed by the Wrongdoer. We want to see whether the initial harm committed by the Wrongdoer gives rise to harmful behavior towards the Scapegoat and whether the identity of the Scapegoat matters.

[Figure 4 around here]

Figure 4 and Table 2 show the main results. First, we document evidence that a substantial fraction of subjects respond to observing the Wrongdoer's misbehavior by lowering the payoff of the Scapegoat (Panel B of Figure 4). Only 4% of subjects diminish the payoff of the Scapegoat when the Wrongdoer did zero harm, but the number increases to 23%-26% for the situations with the Wrongdoer destroyed 4, 6, or 8 euro of Victim's earnings. 23% of subjects decrease the payoff of the Scapegoat by a larger amount in a situation when maximum harm is committed, as compared to when no harm is done by the Wrongdoer. Thus, the effect of observing harm on the prevalence of punishment of Scapegoats is around one third, as compared to punishment of Wrongdoers. In terms of size of punishment, when the Wrongdoer does no harm, we observe only a small reduction of the Scapegoat's payoff (0.12 euro on average), whereas in a situation with maximum harm, it increases to 1.10 euro (Panel A of Figure 4). Again, the punishment

response is weaker, specifically around 20%, when the Scapegoat can be harmed, as compared to when subjects can directly punish the Wrongdoer (0.12 vs. 0.64 euro per 1 euro harm).[19]

[Table 2 around here]

Next, we study whether the ethnic identity of the Scapegoat matters. When no harm is done by the Wrongdoer, there is no evidence of discrimination against the Roma ethnic minority— on average their payoff is lowered by 0.11 euro, while the payoff of the Scapegoat from the majority population is lowered by an average of 0.13 euro. Importantly, when analyzing situations in which the Wrongdoer harmed the Victim, we find a systematic difference in responses between the Scapegoat SAME and Scapegoat OTHER conditions— punishment of Scapegoats is twice as severe when the Scapegoat is from the Roma minority than when the Scapegoat is from the majority population. Specifically, in Scapegoat SAME, an increase in harm intensity by one additional euro motivates Punishers to lower the Scapegoat's earnings by an additional 0.08 euro. In Scapegoat OTHER, the effect doubles to 0.16 euro, and the difference between SAME and OTHER is statistically significant at the 1% level (p=0.002, Column 1 of Table 2). Due to such magnified punishment of the Scapegoat in OTHER, discrimination against the ethnic minority gradually rises with greater harm intensity and becomes statistically significant for situations when the Wrongdoer destroyed 4, 6 or 8 euros of the Victim's earnings (Figure 5, p-values = 0.008, 0.002, 0.011, respectively).

[Figure 5 around here]

The interaction effect of the Wrongdoer's level of harm committed and OTHER on punishment of the Scapegoat is driven by the extensive as well as the intensive margin (Columns 2 and 3 of Table 2), and in both cases it is statistically significant at the 5% level. Each additional euro destroyed by the Wrongdoer leads to an increase in the proportion of those who decide to punish the Scapegoat, by 2 percentage points in Scapegoat SAME and by 3 percentage points in Scapegoat OTHER. Among those who decide to

[19] As expected, punishment of the Wrongdoer and punishment of the Scapegoat are strongly positively correlated. Specifically, for the maximum harm committed by the Wrongdoer, 21% of Punishers do not punish anyone, 54% punish only the Wrongdoer, 23% punish both the Wrongdoer and the Scapegoat, while just 2% punish only the Scapegoat.

scapegoat, the amount destroyed from the Scapegoat's earnings increases by 0.20 euro in Scapegoat SAME and by 0.39 euro in Scapegoat OTHER.

The results are robust to controlling for various additional variables (Table A4). Columns 1-7 document that the coefficients for harm intensity and for the interaction between harm intensity and OTHER hardly change at all when we control for design features, including experimenter fixed effects, the subject's observable characteristics, location fixed effects and parental education level. The coefficients remain statistically significant at the 1% level, and the magnitude stays the same. The results are similar if we use a non-linear specification for the harm caused by the Wrongdoer (Columns 1-2 of Table A5).

Finally, we study heterogeneity and explore whether the observed biases in punishment, both of the Wrongdoer and of the Scapegoat, are driven by an easily identifiable subgroup of individuals. We focus on three characteristics of majority Punishers, for which there is enough variation in our sample, and divide the sample based on gender, student status (secondary school vs. university students) and parental education (as a proxy for socio-economic status). Overall, we do not find evidence that a specific group of subjects drives the results (Tables A6 and A7). The results are qualitatively similar for male and female subjects (although for punishment of the Scapegoat the interaction effect is primarily driven by men), across parental education and, interestingly, they also hold for university students. This suggests that even the future 'elites' are not immune to group biases in punishment.

To summarize, members of the minority group receive systematically harsher punishment than members of the majority group. This is the case when the punished individuals are clearly responsible for the wrongdoing, but also in situations when punished individuals did not contribute to the wrongdoing. The results are consistent with minority scapegoating -- spectators channeling their anger triggered by an observed injustice onto the minority group. In the following subsection, we consider potential alternative explanations for why the majority group engages in magnified punishment of the Roma-minority members.

3.3 Towards understanding mechanisms behind shifting of punishment on minorities

In this sub-section, we discuss and present additional results that speak to the underlying mechanisms of the main effects. We argue that the results support the interpretation that the observed greater inclination to shift punishment on the Roma minority originates in a stronger preference to negatively reciprocate, in terms of both retaliation against Wrongdoer as well as in terms of harming an unrelated party. We also show that the main effects are unlikely to be driven by statistical discrimination, over-attribution of responsibility, or due to "punishment by association" (collective punishment). Finally, we discuss whether shifting of punishment on an out-group is specific to behavior of a dominant group towards ethnic minorities, or whether it extends to any interactions across ethnic boundaries. We also consider whether the effects can be explained by differences in perceptions about socio-economic status of Roma and non-Roma.

3.3.1 Why is shifting of punishment larger when the Roma minority is a Scapegoat?

Statistical discrimination and over-attribution of responsibility. By virtue of the experimental design, beliefs and statistical discrimination are unlikely to explain our findings. First, subjects were randomly allocated into SAME and OTHER and were provided the same information about the nature and the extent of the harm before making their punishment decisions. This essentially rules out the possibility that greater punishment of Roma minority members is because subjects thought that Roma Wrongdoers committed more serious harm than Wrongdoers from the majority group in our experiment. Also, we directly elicited beliefs about the destructiveness of Wrongdoers in SAME and in OTHER conditions, and do not find evidence supporting the idea that subjects believed Roma Wrongdoers to be more likely to harm Victims than majority Wrongdoers.[20]

[20] Specifically, we asked subjects which of the five possible actions they think the Wrongdoer who was matched with them took. Panel A of Table A8 shows the distributions of beliefs, which are not significantly different for the Wrongdoer from the majority population and for the Wrongdoer from the Roma ethnic minority. Most subjects believe that the Wrongdoer did not commit any harm (61% in SAME and 64% in OTHER). Yet, a non-negligible number believe that the Wrongdoer destroyed some part of the earnings of the Victim and 6% of participants in SAME and 8% in OTHER believe that maximum harm was committed. Furthermore, to measure the beliefs in a more detailed way for each individual, we asked the participants to report their beliefs about how many of the twenty potential Wrongdoers displayed in the picture chose each of the five actions. On average, subjects believe that slightly more than eight Wrongdoers decided not to do any harm to the Victim, while around three out of twenty decided to destroy

Moreover, the observed bias in punishment is unlikely to be driven by differences in beliefs about future interactions – for example, by the possibility that Punishers would punish the majority Wrongdoers (or Scapegoats) less due to greater fear that they would face revenge from them, either within or outside of the (field) lab. Punishers knew Wrongdoers and Scapegoats would not have any opportunity to take revenge after their punishment decision, the interactions were one-shot and anonymous, with Wrongdoers and Scapegoats coming from different localities.

Finally, Punishers faced no uncertainty about who was responsible for the wrongdoing, since the experimental protocol made it clear that Wrongdoers caused the harm to the Victim, while Scapegoats did not. Thus, shifting of punishment in our experiment is unlikely to arise due to over-attribution of responsibility, and subsequent blaming of minority groups for injustices done to the majority group members when the source of harm is unclear.

Collective responsibility. A plausible mechanism for why people may punish innocent individuals from other ethnic groups is the notion of collective responsibility. "Punishment by association" is described as retaliation directed not only against the wrongdoer, but also against other members of his group who have no direct association with the perpetrator or direct control over his actions (Lickel, Schmader, and Hamilton 2003; Cushman, Durwin, and Lively 2012). Thus, in our experiment, collective responsibility would predict that punishment of minority Scapegoats is driven specifically by the condition in which the Wrongdoer was also from the minority group. To test this, we take advantage of the orthogonal experimental variation of identity of the Wrongdoer and of the Scapegoat.

Figure 6 displays the results graphically for all four possible combinations of ethnic identity of the Wrongdoer and of the Scapegoat, where the outcome of interest is punishment of the Scapegoat. In all four

all the Victim's earnings (Panel B of Table A8). The distributions are similar across the Wrongdoer SAME and OTHER condition, except that more OTHER (Roma) counterparts are believed to have committed the maximum harm (2.6 vs. 3.1 out of twenty Wrongdoers).

situations, subjects respond to greater harm done with greater punishment of the Scapegoat. This response is, however, not stronger in situations when the Scapegoat has a shared ethnicity with the Wrongdoer (solid lines).

[Figure 6 around here]

Regression analysis tests this more formally. First, we restrict the sample to subjects who were informed that the Wrongdoer was from the majority group. Thus, magnified punishment of the minority Scapegoat cannot be explained by the desire to punish collectively based on the shared ethnicity of Wrongdoer and Scapegoat. Yet, we still find a strong bias in punishment, as indicated by the positive and statistically significant coefficient for the interaction term between harm intensity inflicted by the majority Wrongdoer and Scapegoat OTHER condition ($p=0.006$, Column 1 of Table 3).[21] Further, we turn to a situation when the Wrongdoer is from the Roma minority, and thus a setting in which we would expect to observe intensified punishment of Roma Scapegoats if collective punishment were the driving mechanism (Column 2). The coefficient for the interaction term between harm intensity and Scapegoat OTHER is positive but not statistically significant ($p=0.138$), and the magnitude is somewhat (though not significantly) smaller than when the Wrongdoer is from the majority group. Furthermore, in Column 4, we hold the Roma identity of the Scapegoat constant and estimate whether scapegoating becomes stronger in a situation where the Wrongdoer is also from the Roma minority, as compared to when the Wrongdoer is from the majority group. We do not find evidence for this—when Scapegoats are OTHER, the coefficient for the interaction term between harm intensity and Wrongdoer OTHER is in fact negative and not statistically significant.[22]

[Table 3 around here]

[21] This result also indicates that our findings on differential treatment of minority Scapegoat is unlikely to be explained by pure conformism – Punishers simply copying the behavior of Wrongdoers. Assuming that conformism increases with a similarity of the decision, this explanation would imply larger punishment of majority as compared to minority Scapegoats, if both the Wrongdoer and Victim are from the majority group. This is an opposite pattern of what we find.

[22] In Column 3 of Table 3, we study the role of collective responsibility for the Scapegoat from the majority population. In this case, we also find no evidence that collective responsibility plays a role, since the sensitivity to harm intensity is not greater if the Wrongdoer is of the same ethnicity as the Scapegoat, i.e. from the majority population.

Altogether, this set of results does not support the interpretation that the observed magnified punishment of Scapegoats from the Roma minority is driven by collective responsibility.[23] In fact, we find that the minority Scapegoats are more harshly punished regardless of the ethnic identity of the Wrongdoer. In other words, the minority is punished for any injustice done to the member of the majority group, *including those that originate from within the majority group*.

Retaliatory preferences. Plain unconditional outgroup hate, a strong form of taste-based discrimination, implies that decision-makers may treat the ethnic minority systematically more harshly than members of their own group, independently of the social context. Thus, differential treatment of the discriminated group should be relatively stable. Yet, we find that this is not the case, since we observe that when there is no harm committed against a Victim, and thus no injustice, there is no evidence of discrimination against the minority. Discrimination against the Roma minority is *activated* when subjects observe harm committed against a member of their own group and have an opportunity to punish an unrelated party or the Wrongdoer.

Since the emergence of discriminatory behavior is conditional on observing harm against a member of own group, the results support the interpretation that the biases in punishment against the Roma minority originate in a magnified preference to retaliate. An open question beyond the scope of this paper is about the psychological "micro-foundation" of such elevation in retaliatory preference when interacting with Roma minority. One interesting possibility is that some people harbor latent discriminatory preferences, which are kept at bay during normal emotional states, or when there is no scope for excuses for engaging in discriminatory behavior. When the same individuals become angered by experiencing or witnessing injustice affecting themselves or people they care about, their ability to control their deep urges may be

[23] Note that, by design, we focus on the pure form of enforcement of collective responsibility, but at the same time we close some of the mechanisms behind collective punishment that might be relevant in real life, in particular the role of instrumental punishment. In our experiment, the Punisher knows that the Scapegoat will not be informed about the Wrongdoer's action (and thus what action triggered his punishment), and that the Wrongdoer will not be informed about any punishment, and thus the Punisher cannot send a moral message to the Wrongdoer by punishing the Scapegoat.

weakened, and latent discriminatory preferences may get on surface. Similarly, observation of injustice may provide an excuse and reduce self-image costs from acting based on a latent animus. Thus, these aspects make decisions whether to engage in a retaliatory behavior particularly discrimination-prone.

3.3.2 Group attribute

Much of the literature on scapegoating highlights that such behavior is specific to decision-makers from the majority group when interacting with members of negatively stereotyped, vulnerable and smaller groups (Allport 1954; Bettelheim and Janowitz 1950), rather than vice versa. To test this idea, we measured the behavior of Roma punishers in the same experimental set up and compared their punishment behavior with punishers from the majority group. We find a qualitatively similar pattern of Wrongdoer punishment as we did for the majority population. First, when no harm is done by the Wrongdoer, we find no evidence of discrimination against the majority population—the extent of punishment of the Wrongdoer from the Roma and majority population is very similar (0.90 and 0.92 euro on average, respectively). Second, harmful actions of the Wrongdoer against a Roma subject triggers magnified revenge towards OTHER (majority) Wrongdoers as compared to SAME (Roma) Wrongdoers. While in SAME, when a Wrongdoer destroys one more euro of the Victim's earnings, the intensity of punishment increases by 0.19 euro, in OTHER it goes up by 0.25 euro. The coefficient for an interaction term between the harm intensity and OTHER is statistically significant at the 10% level (p=0.097, Column 1 of Table A9).[24] Consequently, a gap in the punishment of majority vs. Roma Wrongdoers opens up with greater harm intensity and becomes marginally statistically significant for maximum harm (p-value = 0.099, Figure A3).

[24] While the patterns are qualitatively similar, the strength of the effects and the extent of punishment differ across the samples of decision-makers from the Roma and from the ethnic majority population. First, the extent of punishment in a situation where no harm is committed by the Wrongdoer is higher among Roma than among ethnic majority decision-makers (0.91 vs. 0.24 euro, respectively). Second, the extent of punishment in the situation of maximum harm is lower among Roma than majority decision-makers (2.71 vs. 5.23 euro, respectively). Thus, Roma decision-makers were less sensitive to harm intensity. Specifically, in the SAME condition, each additional euro destroyed by the Wrongdoer leads to an increase in punishment by 0.19 euro by Roma and by 0.56 euro by ethnic majority decision-makers.

Next, we explore the behavior of Roma decision-makers towards the Scapegoat (Figure A4 and Table A10). First, we find that Roma subjects punish the Scapegoat, and the extent of scapegoating increases with harm intensity. In a situation where no harm is done by the Wrongdoer, the subjects reduce on average 0.97 euro of the Scapegoat's payoff, while with maximum harm they reduce the Scapegoat's payoff by 1.72 euro. Importantly, when we compare choices in Scapegoat SAME and Scapegoat OTHER, we do not find evidence of ethnic bias in any of the five situations with different harm intensity (Figure A4). Also, in the regression analysis, the coefficient for the interaction between harm intensity and Scapegoat OTHER is small in magnitude and not statistically significant (p=0.737, Column 1 of Table A10).

The results are robust to controlling for design features, location fixed effects, and the education of the parents (Column 4 of Tables A9 and A10). Among Roma Punishers, incomplete understanding is a potential concern, since only 77% of the subjects answered all comprehension questions correctly. To alleviate this concern, we excluded the 23% of subjects who demonstrated imperfect understanding. The results are similar to the original results (Column 5).

Overall, we find weaker evidence of biases in punishments by the minority group against the majority group, as compared to biases of the majority group against the minority group. This result is supportive of the idea that ethnic minorities are more likely to be the subjects of scapegoating from majority groups than vice versa. Nevertheless, this evidence should be taken only as the first step towards understanding which type of real-life characteristics makes certain groups prone to become subject of scapegoating, because we compare only two groups of Punishers that are different across many dimensions, beyond their ethnic majority and minority status, and more research is needed to explore this aspect.

Next, we aim to gauge whether the observed bias in punishment can be explained by differences in perceived socio-economic status, rather than ethnic minority status per se, since the pictures in SAME and OTHER may signal not only ethnicity, but also the socioeconomic status. To do so, we elicited subjects'

perceptions of the socio-economic status of the individuals in the photos, after completing their experimental choices. Specifically, the participants reported their estimates of the employment status and the father's education of the pictured individuals. As expected, the perceptions of individuals in SAME and OTHER indeed differ widely (Table A11). People in photos signaling Roma minority are significantly more likely to be perceived as unemployed, less likely to be students, and less likely to have educated fathers.

We test whether the coefficient for the interaction effect between harm intensity and OTHER is robust to controlling for perceptions of unemployment and its interaction with harm intensity, and to controlling for perceptions of the father's education and its interaction with extent of harm inflicted. Table A12 shows the results for direct punishment of the Wrongdoer and Table A13 for punishment of the Scapegoat. In both tasks, we find that the estimated interaction effects between harm intensity and OTHER on punishment are similar to the baseline estimates. Also, we observe that Punishers who believe that OTHER Wrongdoers or Scapegoats are from a low SES background are not more likely to punish, compared to Punishers who believe Wrongdoers or Scapegoats are not from low SES, once we control for SAME/OTHER conditions.

These patterns suggest that differences in perceptions of socio-economic status, which are associated with SAME and OTHER conditions, cannot fully explain our findings. That said, note that the allocation to OTHER and SAME conditions is exogenous since it was randomly allocated, while the perceptions of parental education and employment status of the people in the photos are endogenous, and thus this analysis needs to be taken with a grain of salt.

4. Conclusion

Social scientists have long speculated that members of dominant social groups have a tendency to shift punishment for social ills originating within their own group onto innocent members of other, weaker groups. This behavior, termed scapegoating, is sometimes considered to be an important psychological

mechanism in the emergence of pogroms, witch-hunts, and large-scale violence. However, clear experimental evidence documenting the behavioral phenomenon in controlled environments had been missing so far. This is what we provide in this paper. We have developed a new incentivized task, the Punishing the Scapegoat Game, to uncover how observing harmful actions against members of one's own group shapes the punishment of innocent individuals. We study the behavior of young adults in Eastern Slovakia, a region experiencing intergroup tensions. In line with the idea that minorities are often treated as scapegoats by dominant groups, we show that Punishers from the ethnic majority group systematically redirect punishment onto innocent members of the Roma minority group for harmful actions committed by other people. Further, when they can punish actual Wrongdoers, they punish Wrongdoers from the minority group more harshly than Wrongdoers from the majority group, for the same wrongdoing.

This paper leaves several interesting questions open for future research. We study interactions between one majority group and one minority group (Roma people), in one setting in which the minority is largely segregated, economically disadvantaged, and deprecated. Clearly, more research is needed to assess whether these findings are generalizable to other settings and to pin down which specific group attributes make certain individuals or groups convenient scapegoats. In this paper, we also study the influence of observing social ills—someone doing harm to a member of one's own group for no apparent reason. Although beyond the scope of our study, an interesting direction to explore is whether similar aggressive discriminatory responses are activated by experiencing personal ills, such as an income shock or stress. We believe the experimental paradigm illustrated in this paper is a well-suited tool to address some of these questions.

At face value, our findings have several potentially important implications. First, we show that pure observation of injustice and wrongdoing against the individual's own group activates discriminatory preferences, both when treating wrongdoers as well as innocent individuals. This indicates that courts, and other settings in which people make punishment choices, are particularly discrimination-prone environments, in line with evidence of strong biases against minorities in judicial decisions (Shayo and

Zussman 2011; Alesina and Ferrara 2014; Rehavi and Starr 2014). Second, the results suggest that ethnic minorities are at greater risk when social problems and unfair behavior become salient features of the societal environment. So far, economists have typically attributed sudden spikes in aggressive behavior towards weaker groups to changes in economic incentives or beliefs about the likelihood of facing a penalty for aggressive behavior (Blattman and Miguel 2010; Grosfeld, Sakalli, and Zhuravskaya 2019; Miguel 2005), assuming that revealed (anti) social preferences towards other groups are stable. In our experiment, economic incentives are held constant and thus cannot explain the scapegoating behavior observed. Of course, this does not imply that economic incentives do not play an important role in real-life aggression towards minority groups. However, our evidence suggests this may not be the complete picture and strengthens the case for taking seriously "behavioral" channels through which deterioration of the social environment may fuel inter-group conflicts.

Acknowledgements

We thank Alexander Cappelen, Ernst Fehr, Michael Kosfeld, Filip Matějka, Matthias Sutter, Bertil Tungodden and seminar participants at the University of Munich, Max Planck Institute in Bonn, UC Berkeley, University of Edinburgh, Goethe University in Frankfurt, Norwegian School of Economics and the Symposium for Social Cohesion in Ethnically Diverse Societies in Tilburg for helpful comments. The data collection was supported by a grant from the Czech Science Foundation (17-13869S), and by the Slovak Research and Development Agency (APVV-0125-12 and APVV-19-0329). Bauer and Chytilová acknowledge support from ERC-CZ/AV-B (ERC300851901).

The research was approved by the Ethical committee at the Technical University of Košice and by the Director of the Institute of Economic Studies at the Faculty of Social Sciences, Charles University.

References

Abbink, Klaus, Jordi Brandts, Benedikt Herrmann, and Henrik Orzen. 2010. "Intergroup Conflict and Intra-Group Punishment in an Experimental Contest Game." *American Economic Review* 100 (1): 420–47.

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

Alesina, By Alberto, and Eliana La Ferrara. 2014. "A Test of Racial Bias in Capital Sentencing." *American Economic Review* 104 (11): 3397–3433.

Alexander, Marcus, and Fotini Christia. 2011. "Context Modularity of Human Altruism." *Science* 1392 (334): 1392–94.

Alexander, Richard. 1987. *The Biology of Moral Systems*. Aldine de Gruyter, New York.

Allport, Gordon W. 1954. *The Nature of Prejudice*. New York: Addison-Wesley.

Almas, Ingvild, Alexander Cappelen, and Bertil Tungodden. 2019. "Cutthroat Capitalism versus Cuddly Socialism : Are Americans More Meritocratic and Efficiency-Seeking than Scandinavians?" *Journal of Political Economy* forth.

Anderson, Robert Warren, Noel D. Johnson, and Mark Koyama. 2017. "Jewish Persecutions and Weather Shocks: 1100–1800." *Economic Journal* 127 (602): 924–58.

Bardhan, Pranab. 2005. *Scarcity, Conflict and Cooperation*. Cambridge, MA: The MIT Press.

Bartling, Björn, and Urs Fischbacher. 2012. "Shifting the Blame : On Delegation and Responsibility." *Review of Economic Studies* 79 (1): 67–87.

Bartoš, Vojtěch, Michal Bauer, Julie Chytilová, and Filip Matějka. 2016. "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition." *American Economic Review* 106 (6): 1437–75.

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.

Baumgartner, Thomas, Lorenz Goette, Rahel Guegler, and Ernst Fehr. 2012. "The Mentalizing Network Orchestrates the Impact of Parochial Altruism on Social Norm Enforcement." *Human Brain Mapping* 1469 (August 2010): 1452–69.

Berge, Lars Ivar Oppedal, Kjetil Bjorvatn, Simon Galle, Edward Miguel, Daniel Posner, Bertil Tungodden, and Kelly Zhang. 2020. "Ethnically Biased? Experimental Evidence from Kenya." *Journal of the European Economic Association* 18 (1): 134–64.

Bernhard, Helen, Urs Fischbacher, and Ernst Fehr. 2006. "Parochial Altruism in Humans." *Nature* 442 (7105): 912–915.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94 (4): 991–1013.

Bettelheim, B., and M. Janowitz. 1950. *Dynamics of Prejudice: A Sociological Study of Veterans*. New York: Harper and Brothers.

Blattman, C., and E. Miguel. 2010. "Civil War." *Journal of Economic Literature* 48 (1): 3–57.

Bolton, Gary, Elena Katok, and Axel Ockenfels. 2005. “Cooperation among Strangers with Limited Information about Reputation.” *Journal of Public Economics* 89 (8): 1457–68.

Cappelen, Alexander W, Astri Drange Hole, Erik Sorensen, and Bertil Tungodden. 2007. “The Pluralism of Fairness Ideals : An Experimental Approach.” *American Economic Review* 97 (3): 818–27.

Chen, Y., and S. X Li. 2009. “Group Identity and Social Preferences.” *American Economic Review* 99 (1): 431–457.

Council of Europe. 2012. *Human Rights of Roma and Travellers in Europe*. Strasbourg: Council of Europe Publishing.

Crossman, Ashley. 2018. “Definition of Scapegoat, Scapegoating, and Scapegoat Theory,” 2018.

Cushman, Fiery, A J Durwin, and Chaz Lively. 2012. “Revenge without Responsibility? Judgments about Collective Punishment in Baseball.” *Journal of Experimental Social Psychology* 48 (5): 1106–10.

Doob, Leonard W, Neal E Miller, Orval Hobart Mowrer, Robert Sears, and John Dollard. 1939. *Frustration and Aggression*. New Haven, CT: Yale University Press.

Dufwenberg, Martin, Uri Gneezy, Werner Güth, and Eric van Damme. 2001. “Direct vs Indirect Reciprocity: An Experiment.” *Homo Oeconomicus* 18: 19–30.

Engelmann, Dirk, and Urs Fischbacher. 2009. “Indirect Reciprocity and Strategic Reputation Building in an Experimental Helping Game.” *Games and Economic Behavior* 67 (2): 399–407.

Esteban, Joan-Maria, and Debraj Ray. 2008. “On the Salience of Ethnic Conflict.” *American Economic Review* 98 (5): 2185–2202.

Fearon, James D., and David D Laitin. 2000. “Violence and the Social Construction of Ethnic Identity.” *International Organization* 54 (4): 845–77.

Fehr, Ernst, and Urs Fischbacher. 2004. “Third-Party Punishment and Social Norms.” *Evolution and Human Behavior* 25 (2): 63–87.

Fehr, Ernst, and Simon Gächter. 2002. “Altruistic Punishment in Humans.” *Nature* 415 (6868): 137–40.

Fershtman, C., and U. Gneezy. 2001. “Discrimination in a Segmented Society: An Experimental Approach.” *Quarterly Journal of Economics* 116 (1): 351–377.

Fischbacher, Urs. 2007. “Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments.” *Experimental Economics* 10 (2): 171–77.

Gächter, Simon, Elke Renner, and Martin Sefton. 2008. “The Long-Run Benefits of Punishment.” *Science* 322 (5907): 1510–1510.

Girard, Rene. 1979. *Things Hidden Since the Foundation Ofthe World*. Stanford, CA: Stanford University Press.

Glick, Peter. 2005. “Choice of Scapegoats.” In *On the Nature of Prejudice*, edited by John F Dovidio, Peter Glick, and Laurie Budman, 244–61. Blackwell Publishing.

Goette, Lorenz, David Huffman, and Stephan Meier. 2006. “The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Usng Random Assignment to Real Social Groups.” *American Economic Review Papers and Proceedings* 96 (2): 212–16.

Goette, Lorenz, David Huffman, Stephan Meier, and Matthias Sutter. 2012. “Competition Between Organizational Groups: Its Impact on Altruistic and Antisocial Motivations.” *Management Science*

58 (5): 948–60.

Grosfeld, Irena, Seyhun Orcan Sakalli, and Ekaterina Zhuravskaya. 2019. “Middleman Minorities and Ethnic Violence: Anti-Jewish Pogroms in the Russian Empire.” *Review of Economic Studies* forth.

Henrich, Joseph, Jean Ensminger, Richard Mcelreath, Abigail Barr, Clark Barrett, Alexander Bolyanatz, Juan Camilo Cardenas, Michael Gurven, and Edwins Gwako. 2010. “Markets, Religion, Community Size, and the Evolution of Fairness and Punishment.” *Science* 1480 (March 2010): 1480–84.

Herrmann, Benedikt, Christian Thoni, and Simon Gächter. 2008. “Antisocial Punishment across Societies.” *Science* 319 (5868): 1362–67.

Hjort, Jonas. 2014. “Ethnic Divisions and Production in Firms.” *Quarterly Journal of Economics* 129 (4): 1899–1944.

Horowitz, Donald L. 1985. *Ethnic Groups in Conflict*. London: University of California Press.

Kant, Immanuel. 1965. *The Metaphysical Elements of Justice*. New York: Bobbs-Merill Company.

Levin, Jack, and Jack McDevitt. 2002. *Hate Crimes Revisited: America’s War Against Those Who Are Different*. Basic Books.

Lickel, Brian, Toni Schmader, and David L Hamilton. 2003. “A Case of Collective Responsibility: Who Else Was to Blame for the Columbine High School Shootings?” *Personality and Social Psychology Bulletin* 29 (2): 194–204.

Miguel, Edward. 2005. “Poverty and Witch Killing.” *Review of Economic Studies* 72 (4): 1153–72.

Newman, Leonard S., and Tracy L. Caldwell. 2005. “Alloport’s ‘Living Inkblots’: The Role of Defensive Projection in Stereotyping and Prejudice.” In *On the Nature of Prejudice: Fifty Years after Allport*, edited by John F. Dovidio, Peter Glick, and Laurie Rudman, 377–93. Oxdford: Blackwell Publishing.

Nowak, Martin, and Karl Sigmund. 2005. “Evolution of Indirect Reciprocity.” *Nature* 437 (7063): 1291–98.

Prediger, Sebastian, Bjorn Vollan, and Benedikt Herrman. 2014. “Resource Scarcity and Antisocial Behavior.” *Journal of Public Economics* 119: 1–9.

Rehavi, M. Marit, and Sonja B. Starr. 2014. “Racial Disparity in Federal Criminal Charging and Its Sentencing Consequences.” *Journal of Political Economy* 122 (6): 1320–54.

Schiller, Bastian, Thomas Baumgartner, and Daria Knoch. 2014. “Intergroup Bias in Third-Party Punishment Stems from Both Ingroup Favoritism and Outgroup Discrimination.” *Evolution and Human Behavior* 35: 383–93.

Shayo, Moses, and Asaf Zussman. 2011. “Judicial Ingroup Bias in the Shadow of Terrorism.” *Quarterly Journal of Economics* 126 (3): 1447–84.

Stanca, Luca. 2009. “Measuring Indirect Reciprocity: Whose Back Do We Scratch?” *Journal of Economic Psychology* 30 (2): 190–202.

Staub, Ervin. 1992. *The Roots of Evil: The Origins of Genocide and Other Group Violence*. Cambridge, MA: Cambridge University Press.

Stenner, Karren. 2005. *The Authoritarian Dynamic*. New York: Cambridge University Press.

Stenner, Karren, and Jonathan Haidt. 2018. “Authoritarianism Is Not a Momentary Madnes, but an

Eternal Dynamic with Liberal Democracies." In *Can It Happen Here? Authoritarianism in America*, edited by Cass R. Sunstein, 174–93. New York: HarperCollins.

Tajfel, Henry. 1981. *Human Groups and Social Categories: Studies in Social Psychology*. Cambridge: Cambridge University Press.

Voigtlander, Nico, and Hans-Joachim Voth. 2012. "Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany." *The Quarterly Journal of Economics* 127 (3): 1339–92.

Yanagizawa-Drott, David. 2014. "Propaganda and Conflict: Evidence from the Rwandan Genocide." *Quarterly Journal of Economics* 129 (4): 1947–94.

TABLE 1: PUNISHMENT OF THE WRONGDOER

Dependent variable	Punishment of the Wrongdoer (intensity)	Punishment of the Wrongdoer (yes)	Punishment of the Wrongdoer (intensity)	Punishment of the Wrongdoer (intensity)
Sample	All	All	Punishment of the Wrongdoer = yes	All
	(1)	(2)	(3)	(4)
Harm intensity	0.56***	0.07***	0.63***	0.56***
	(0.03)	(0.00)	(0.04)	(0.04)
Wrongdoer OTHER	-0.26**	-0.03	-0.53**	-0.32**
	(0.11)	(0.03)	(0.23)	(0.13)
Harm intensity*Wrongdoer OTHER	0.14***	0.02***	0.14***	0.14***
	(0.05)	(0.01)	(0.05)	(0.05)
Controls	baseline	baseline	baseline	full
Mean baseline (Wrongdoer SAME, 0 harm)	0.35	0.10	3.41	0.35
Observations	1,685	1,685	1,044	1,685
R-squared	0.417	0.201	0.540	0.453

Notes: OLS, standard errors (in parentheses) are clustered at the Punisher level. *** denotes p<0.01, ** p<0.05, and * p<0.1. The dependent variable in Columns 1,3,4 is the extent of punishment of the Wrongdoer (EUR 0-8). In Column 2, the dependent variable indicates that the Punisher chose non-zero punishment of the Wrongdoer. "Harm intensity" captures the harm caused by the Wrongdoer to the Victim (EUR 0-8). Punishers (and Victims) are from the majority ethnic group. "Wrongdoer OTHER" indicates that the Wrongdoer comes from a different ethnic group (Roma minority) than the Punisher. Baseline controls include the gender and age of the Punisher, and a dummy variable indicating that the Punishing the Wrongdoer Game took place before the Punishing the Scapegoat Game. Full controls also include a dummy variable "Scapegoat OTHER" (equal to one if the Scapegoat comes from a different ethnic group than the Punisher), experimenter fixed effects, a dummy variable indicating that the Punisher is a university student (vs. a secondary school student), location fixed effects, education of parents (dummy variables for mother/father with a university degree, dummy variables for education unknown), and a dummy variable indicating that the subject answered all control questions correctly on the first or second attempt. The sample is composed of Punishers from the majority group.

TABLE 2: PUNISHMENT OF THE SCAPEGOAT

Dependent variable	Punishment of the Scapegoat (intensity)	Punishment of the Scapegoat (yes)	Punishment of the Scapegoat (intensity)	Punishment of the Scapegoat (intensity)
Sample	All	All	Punishment of the Scapegoat = yes	All
	(1)	(2)	(3)	(4)
Harm intensity	0.08***	0.02***	0.20***	0.08***
	(0.02)	(0.00)	(0.07)	(0.02)
Scapegoat OTHER	-0.00	0.03	-0.57	-0.04
	(0.08)	(0.03)	(0.37)	(0.09)
Harm intensity*Scapegoat OTHER	0.09***	0.01**	0.19**	0.09***
	(0.03)	(0.01)	(0.08)	(0.03)
Wrongdoer OTHER	-0.14	-0.03	-0.32	-0.14
	(0.13)	(0.03)	(0.24)	(0.13)
Controls	baseline	baseline	baseline	full
Mean baseline (Scapegoat SAME, 0 harm)	0.13	0.04	3.14	0.13
Observations	1,685	1,685	329	1,685
R-squared	0.072	0.050	0.240	0.133

Notes: OLS, standard errors (in parentheses) are clustered at the Punisher level. *** denotes p<0.01, ** p<0.05, and * p<0.1. The dependent variable in Columns 1,3,4 is the extent of punishment of the Scapegoat (EUR 0-8). In Column 2, the dependent variable indicates that the Punisher chose non-zero punishment of the Scapegoat. "Harm intensity" captures the harm caused by the Wrongdoer to the Victim (EUR 0-8). Punishers (and Victims) are from the majority ethnic group. "Scapegoat OTHER" indicates that the Scapegoat comes from a different ethnic group (Roma minority) than the Punisher. "Wrongdoer OTHER" indicates that the Wrongdoer comes from a different ethnic group than the Punisher. Baseline and full controls are defined in Table 1. The sample is composed of Punishers from the majority group.

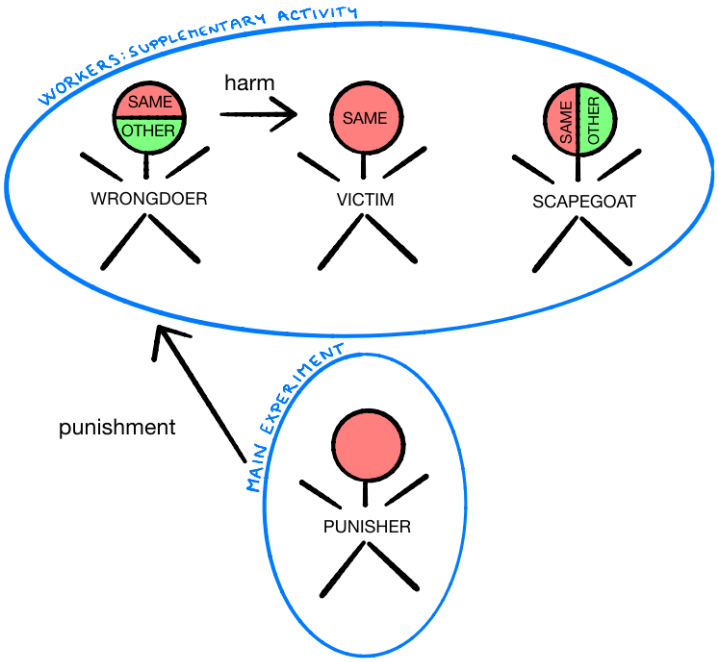
TABLE 3: TESTING FOR COLLECTIVE PUNISHMENT

Dependent variable	Punishment of the Scapegoat (intensity)			
	Wrongdoer	Wrongdoer	Scapegoat	Scapegoat
Sample	SAME	OTHER	SAME	OTHER
	(1)	(2)	(3)	(4)
Harm intensity	0.07**	0.09***	0.09***	0.19***
	(0.03)	(0.02)	(0.02)	(0.03)
Scapegoat OTHER	0.06	-0.07		
	(0.14)	(0.08)		
Harm intensity*Scapegoat OTHER	0.12***	0.06		
	(0.04)	(0.04)		
Wrongdoer SAME			0.02	
			(0.13)	
Harm intensity*Wrongdoer SAME			-0.02	
			(0.04)	
Wrongdoer OTHER				-0.15
				(0.09)
Harm intensity*Wrongdoer OTHER				-0.04
				(0.05)
Controls	baseline	baseline	baseline	baseline
Mean baseline	0.19	0.07	0.07	0.14
Observations	835	850	840	845
R-squared	0.101	0.061	0.036	0.084

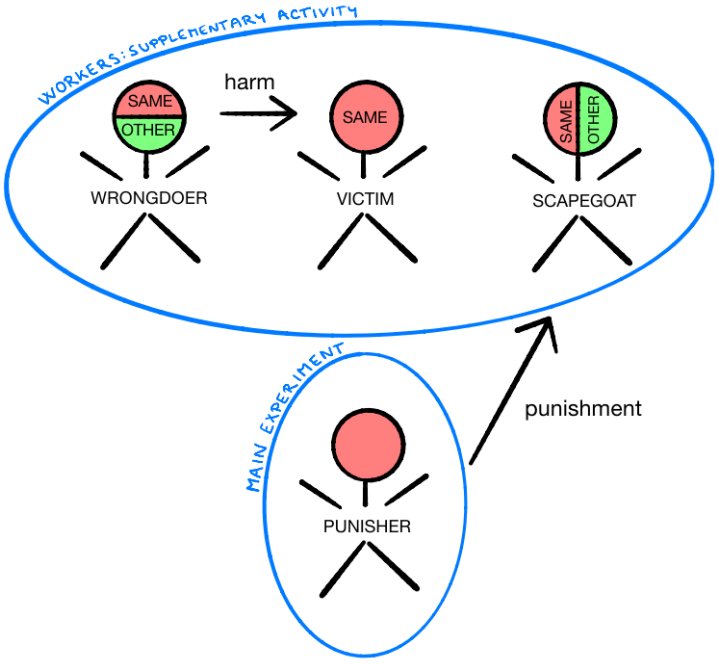
Notes: OLS, standard errors (in parentheses) are clustered at a Punisher level. *** denotes p<0.01, ** p<0.05, and * p<0.1. The dependent variable is the extent of punishment of the Scapegoat (EUR 0-8). "Harm intensity" captures the harm caused by the Wrongdoer to the Victim (EUR 0-8). Punishers (and Victims) are from the majority ethnic group. "Wrongdoer SAME" and "Wrongdoer OTHER" indicate that the Wrongdoer comes from the same ethnic group or a different ethnic group (Roma minority) than the Punisher, respectively. "Scapegoat SAME" and "Scapegoat OTHER" indicate that the Scapegoat comes from the same ethnic group or a different ethnic group (Roma minority) than the Punisher, respectively. Baseline controls are defined in Table 1. The sample is composed of Punishers from the majority group.

FIGURE 1: ILLUSTRATION OF THE EXPERIMENTAL TASKS

(A) PUNISHING THE WRONGDOER GAME



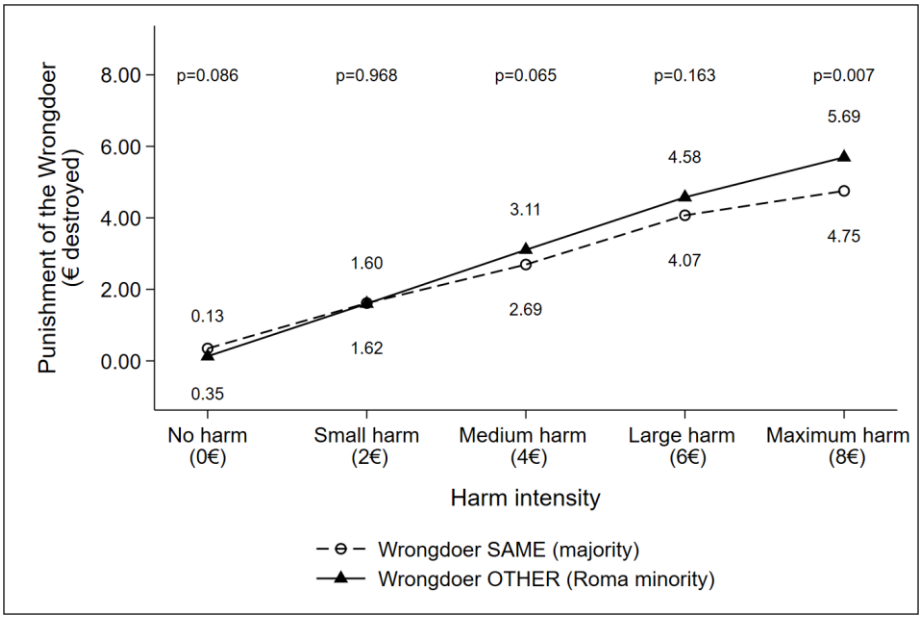
(B) PUNISHING THE SCAPEGOAT GAME



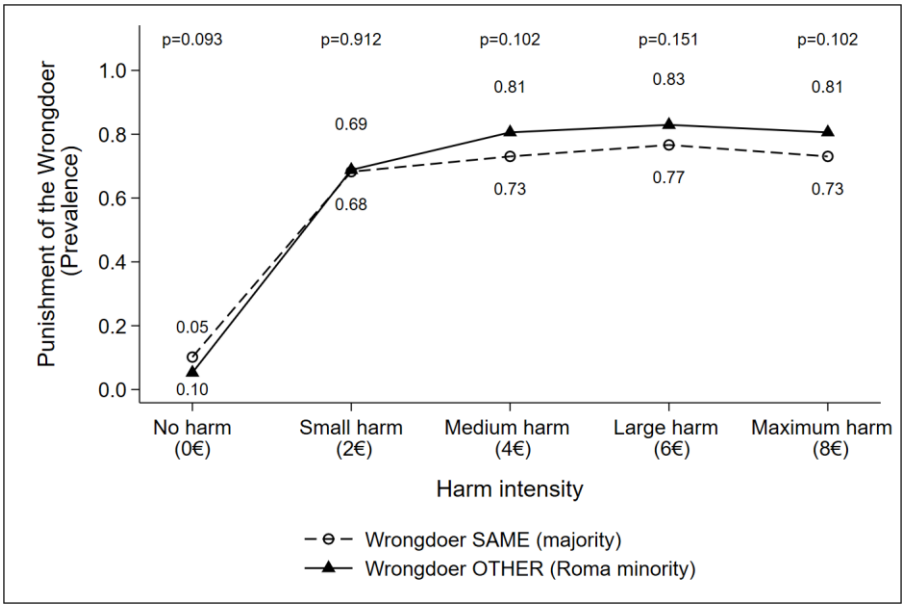
Notes: The main sample are Punishers, who come either from the majority group or from the Roma ethnic minority. Punishers learn about the harm committed by the Wrongdoers towards the Victims and can punish the Wrongdoer (Panel A) or can only punish an innocent bystander—the Scapegoat (Panel B). The ethnic identity of the Wrongdoer and the Scapegoat is manipulated orthogonally: "SAME" indicates that the player (Wrongdoer/Victim/Scapegoat) is of the same ethnicity as the Punisher, while "OTHER" indicates that he is from the other ethnic group.

FIGURE 2: PUNISHMENT OF THE WRONGDOER, BY WRONGDOER'S ETHNICITY (MAJORITY SAMPLE)

(A) INTENSITY OF PUNISHMENT

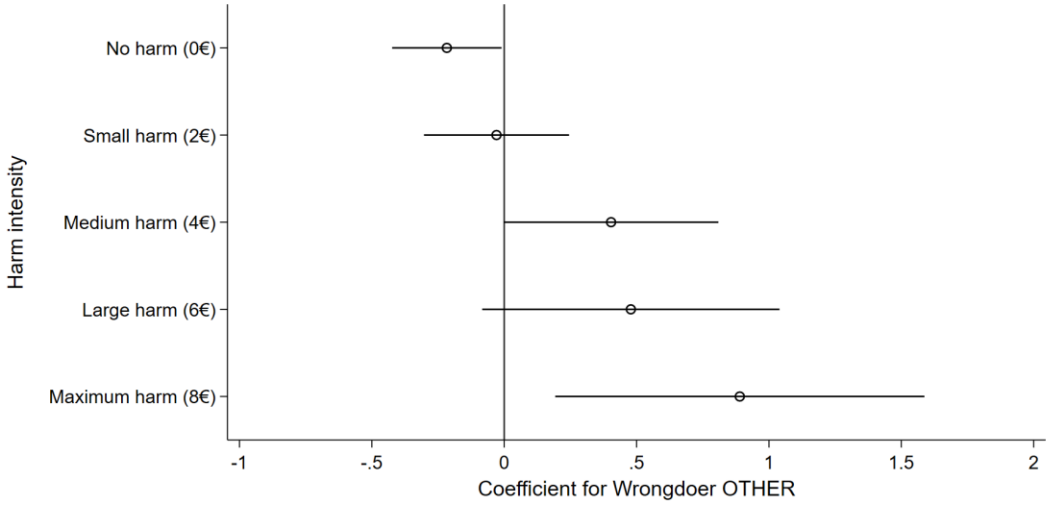


(B) PREVALENCE OF PUNISHMENT



Notes: Mean punishment of the Wrongdoer (Panel A) and the share of Punishers who choose non-zero punishment of the Wrongdoer (Panel B), by the ethnicity of the Wrongdoer and the harm caused by the Wrongdoer to the Victim. Punishers (and Victims) are from the majority ethnic group. "Wrongdoer SAME" indicates that the Wrongdoer also comes from the majority ethnic group, while "Wrongdoer OTHER" indicates that the Wrongdoer is ethnic Roma. Differences between the conditions are tested using the Wilcoxon rank-sum test; p-values are presented at the top. The sample is composed of Punishers from the majority group.

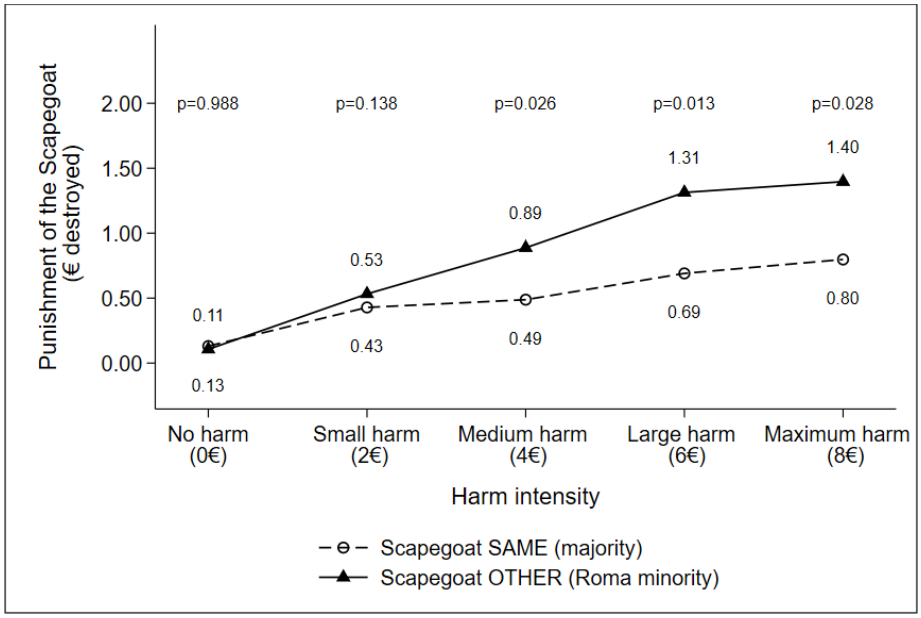
FIGURE 3: DISCRIMINATION AGAINST OTHER WRONGDOERS (MAJORITY SAMPLE)



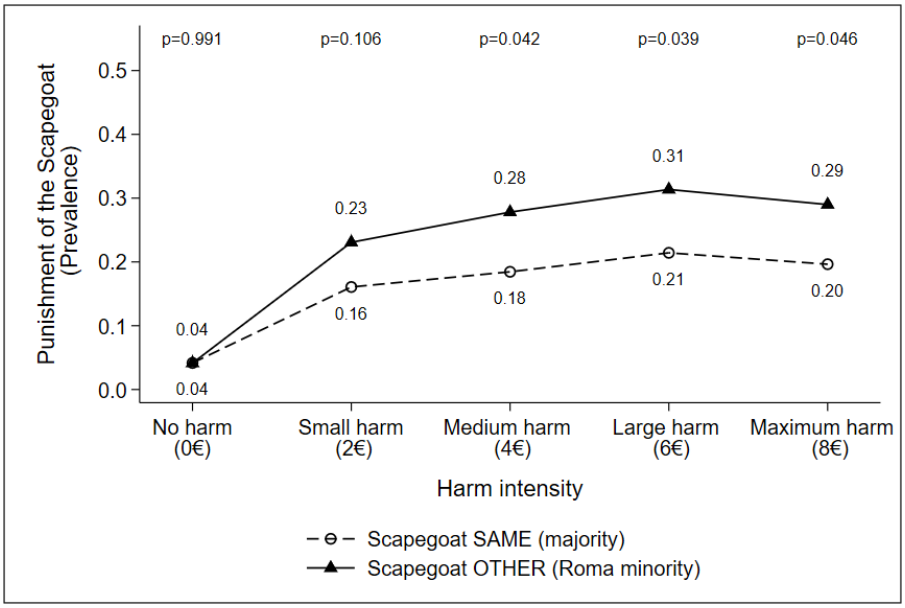
Notes: Estimated coefficients for "Wrongdoer OTHER", with 95% confidence intervals. The dependent variable is the extent of punishment of the Wrongdoer (EUR 0-8). The coefficients are estimated by separate OLS regressions for the five possible levels of harm caused by the Wrongdoer to the Victim. We control for the gender and age of the Punisher, and for a dummy variable indicating that the Punishing the Wrongdoer Game took place before the Punishing the Scapegoat Game. Punishers (and Victims) are from the majority ethnic group. "Wrongdoer OTHER" indicates that the Wrongdoer comes from a different ethnic group (Roma minority) than the Punisher. The sample is composed of Punishers from the majority group.

FIGURE 4: PUNISHMENT OF THE SCAPEGOAT, BY SCAPEGOAT'S ETHNICITY

(A) INTENSITY OF PUNISHMENT

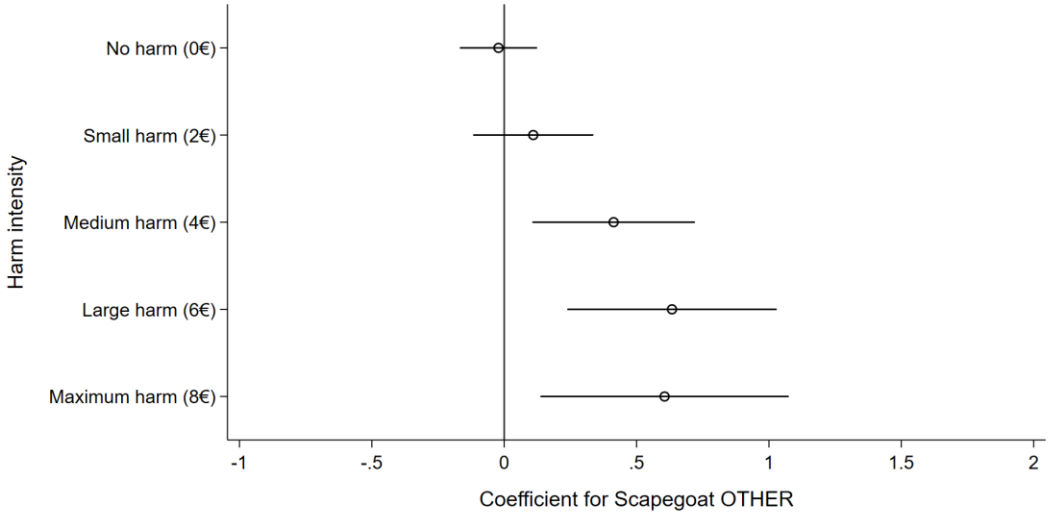


(B) PREVALENCE OF PUNISHMENT



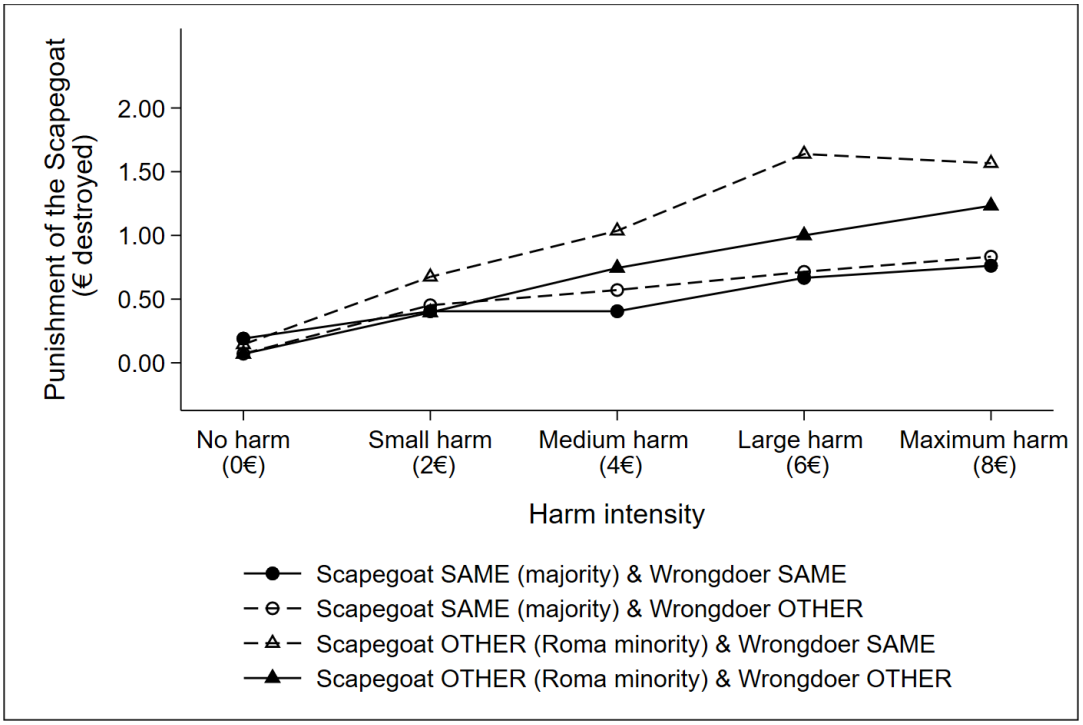
Notes: Mean punishment of the Scapegoat (Panel A) and the share of Punishers who choose non-zero punishment of the Scapegoat (Panel B), by the ethnicity of the Scapegoat and the harm caused by the Wrongdoer to the Victim. Punishers (and Victims) are from the majority ethnic group. "Scapegoat SAME" indicates that the Scapegoat also comes from the majority ethnic group, while "Scapegoat OTHER" indicates that Scapegoat is ethnic Roma. Differences between the conditions are tested using the Wilcoxon rank-sum test; p-values are presented at the top. The sample is composed of Punishers from the majority group.

FIGURE 5: DISCRIMINATION AGAINST OTHER SCAPEGOATS (MAJORITY SAMPLE)



Notes: Estimated coefficients for "Scapegoat OTHER", with 95% confidence intervals. The dependent variable is the extent of punishment of the Scapegoat (EUR 0-8). The coefficients are estimated by separate OLS regressions for the five possible levels of harm caused by the Wrongdoer to the Victim. We control for the gender and age of the Punisher, and for a dummy variable indicating that the Punishing the Wrongdoer Game took place before the Punishing the Scapegoat Game. Punishers (and Victims) are from the majority ethnic group. "Scapegoat OTHER" indicates that the Wrongdoer comes from a different ethnic group (Roma minority) than the Punisher. The sample is composed of Punishers from the majority group.

FIGURE 6: PUNISHMENT OF THE SCAPEGOAT, BY WRONGDOER'S AND SCAPEGOAT'S ETHNICITY (MAJORITY SAMPLE)



Notes: Mean punishment of the Scapegoat, by the ethnicity of the Scapegoat and the Wrongdoer and by the extent of harm caused by the Wrongdoer to the Victim. Punishers (and Victims) are from the majority ethnic group. "Scapegoat SAME" and "Wrongdoer SAME" indicate that the Scapegoat/the Wrongdoer also comes from the majority ethnic group, while "Scapegoat OTHER" and "Wrongdoer OTHER" indicate that the Scapegoat/the Wrongdoer is ethnic Roma. The sample is composed of Punishers from the majority group.

[FOR ONLINE PUBLICATION]

Scapegoating: Experimental Evidence

Michal Bauer, Jana Cahlíková, Julie Chytilová, Gérard Roland, Tomáš Želinský

Supplementary Online Appendix A

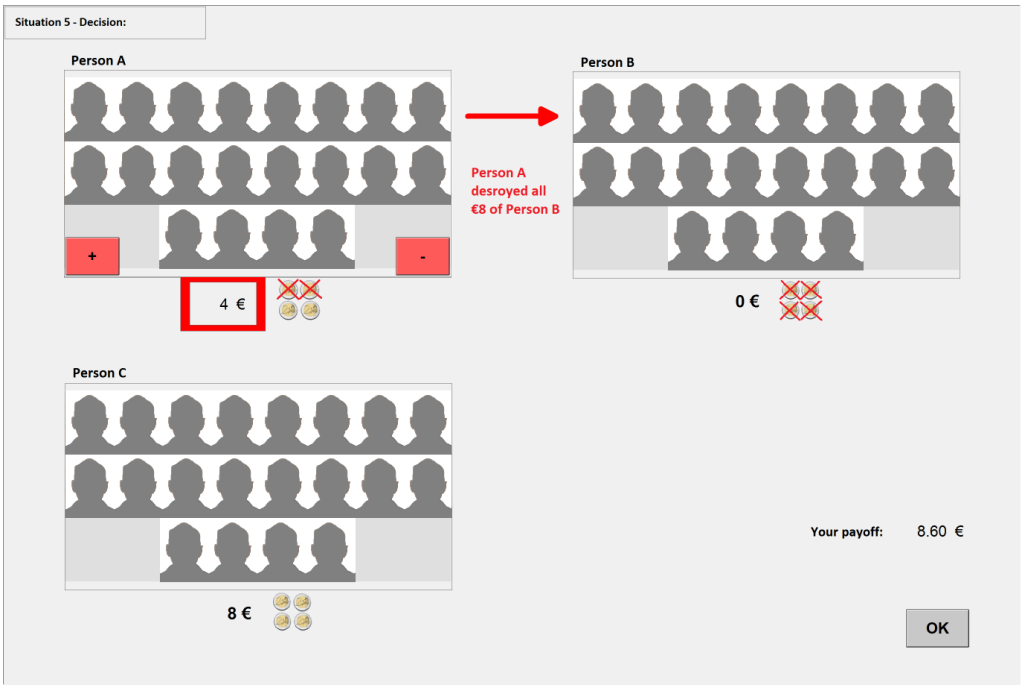
This file contains appendix figure and tables

Appendix Figures A1-A4

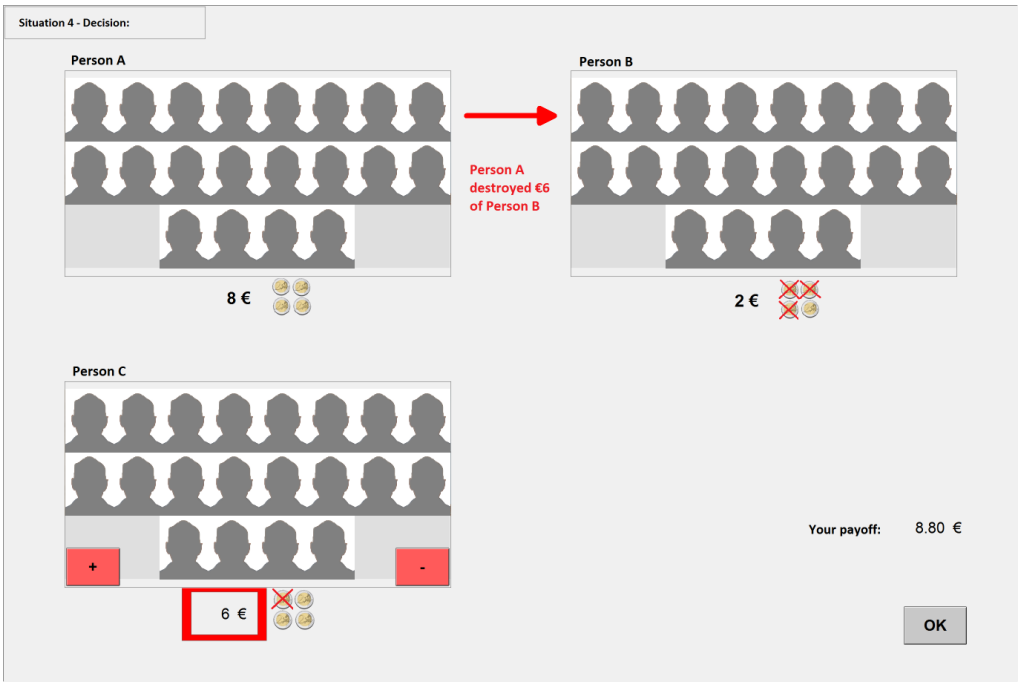
Appendix Tables A1-A13

FIGURE A1: EXAMPLES OF DECISION SCREENS

Panel A: Punishing the Wrongdoer Game



Panel B: Punishing the Scapegoat Game



Notes: Panel A shows the decision screen for the Punishing the Wrongdoer Game, for the situation in which the Wrongdoer (Person A) destroyed all €8 of the earnings of the Victim (Person B). In this example, the Punisher is choosing to reduce the earnings of the Wrongdoer by €4. Panel B shows the decision screen for the Punishing the Scapegoat Game, for the situation in which the Wrongdoer (Person A) destroyed €6 of the Victim's earnings (Person B). In this example, the Punisher is choosing to reduce the earnings of the Scapegoat (Person C) by €2. The default/starting point in all situations is always Person A and Person C getting all €8. In the experiment, each picture shows 20 different passport-style photographs, which vary based on the condition. For privacy reasons, we are not allowed to display the actual pictures of all study participants in publicly available materials.

FIGURE A2: HISTOGRAM OF THE HARM IMPOSED ON THE VICTIMS BY THE WRONGDOERS

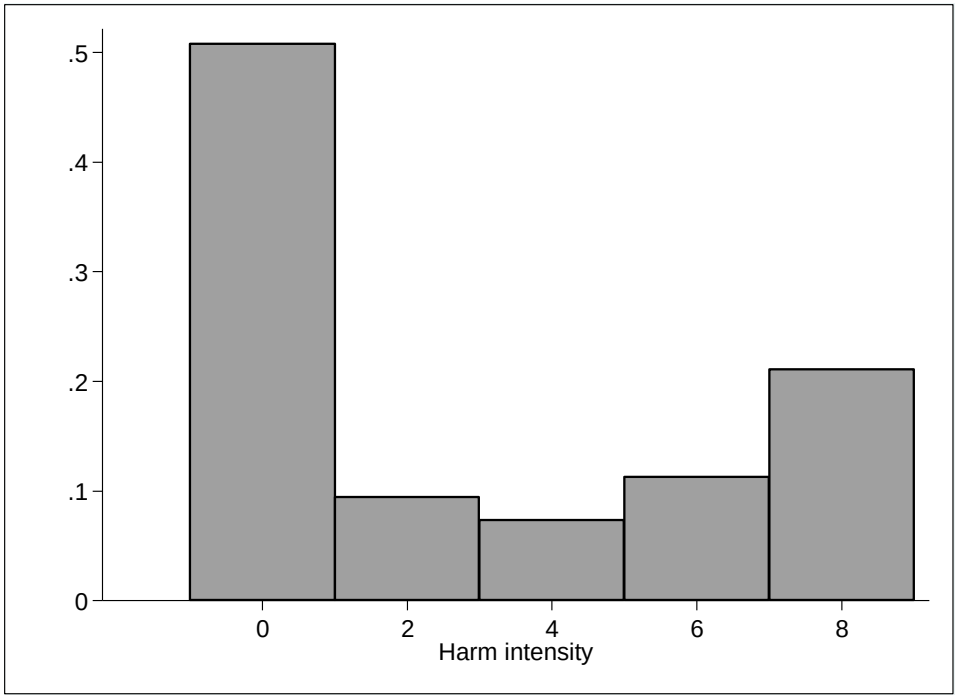
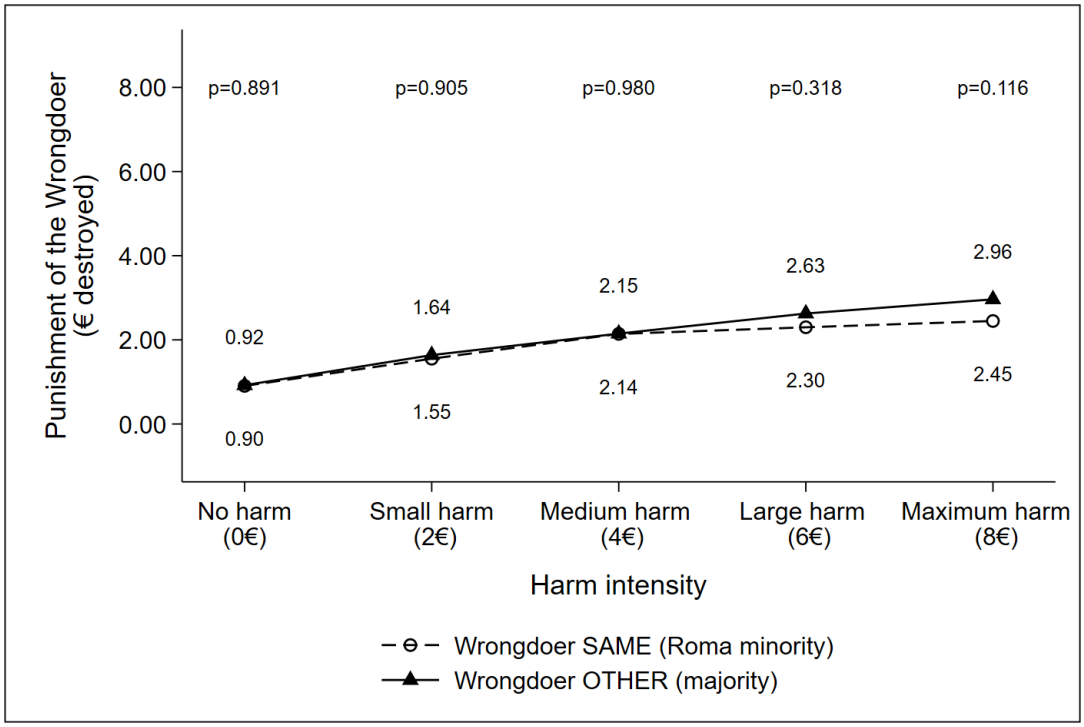
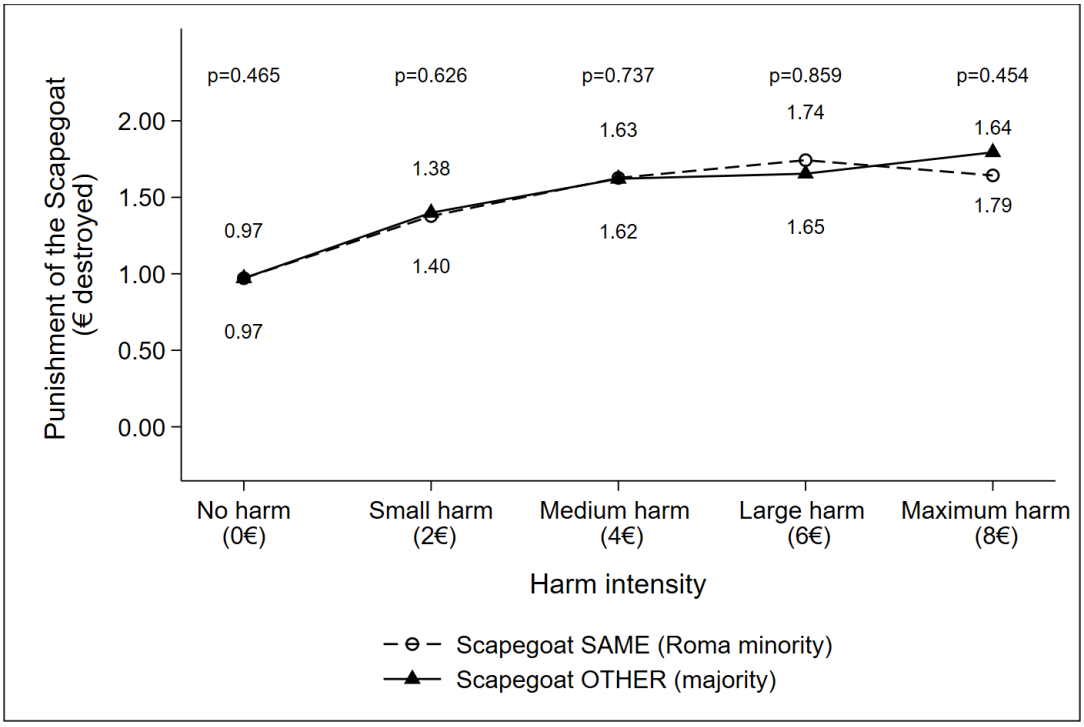


FIGURE A3: PUNISHMENT OF THE WRONGDOER, BY WRONGDOER'S ETHNICITY (ROMA MINORITY SAMPLE)



Notes: Mean punishment of the Wrongdoer, by the ethnicity of the Wrongdoer and the harm caused by the Wrongdoer to the Victim. Punishers (and Victims) are from the Roma ethnic group. "Wrongdoer SAME" indicates that the Wrongdoer also comes from the Roma ethnic group, while "Wrongdoer OTHER" indicates that Wrongdoer has majority ethnicity. Differences between the conditions are tested using the Wilcoxon rank-sum test; p-values are presented on top.

FIGURE A4: PUNISHMENT OF THE SCAPEGOAT, BY SCAPEGOAT'S ETHNICITY (ROMA MINORITY SAMPLE)



Notes: Mean punishment of the Scapegoat, by the ethnicity of the Scapegoat and the harm caused by the Wrongdoer to the Victim. Punishers (and Victims) are from the Roma ethnic group. "Scapegoat SAME" indicates that the Scapegoat also comes from the Roma ethnic group, while "Scapegoat OTHER" indicates that Scapegoat has majority ethnicity. Differences between the conditions are tested using the Wilcoxon rank-sum test; p-values are presented on top.

TABLE A1: DESCRIPTIVE STATISTICS AND RANDOMIZATION CHECKS (MAJORITY SAMPLE)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Means across treatments				
Sample	Whole sample	N	Wrongdoer SAME	Wrongdoer OTHER	Scapegoat SAME	Scapegoat OTHER	Diff (F-test)
Female	0.42	337	0.45	0.39	0.45	0.39	0.31
Age	19.31	337	19.30	19.33	19.35	19.28	0.91
Student (high school)	0.57	337	0.57	0.56	0.57	0.57	0.97
Student (university)	0.43	337	0.43	0.44	0.43	0.43	0.97
Primary school degree	0.54	337	0.54	0.53	0.55	0.53	0.89
Secondary school degree	0.40	337	0.38	0.41	0.38	0.42	0.61
University degree	0.07	337	0.07	0.06	0.08	0.05	0.60
Child	0.00	337	0.01	0.00	0.01	0.00	0.37
Household size	4.24	337	4.32	4.17	4.17	4.31	0.31
Regularly goes to church	0.32	337	0.35	0.29	0.33	0.31	0.48
Father with secondary school	0.74	330	0.75	0.73	0.77	0.72	0.57
Father with university	0.25	330	0.25	0.26	0.23	0.28	0.66
Mother with secondary school	0.68	336	0.69	0.66	0.69	0.66	0.65
Mother with university	0.32	336	0.30	0.34	0.30	0.33	0.64
Father employed	0.96	309	0.96	0.97	0.97	0.96	0.65
Mother employed	0.88	320	0.87	0.89	0.88	0.88	0.81
Family owns a car	0.91	337	0.93	0.89	0.90	0.92	0.31
Family owns a computer	0.99	336	0.99	1.00	1.00	0.99	0.37
Family owns a smartphone	0.96	337	0.96	0.96	0.96	0.96	0.94
Family owns a tablet	0.64	337	0.67	0.60	0.71	0.56	0.01
Control questions first attempt (max=6)	5.93	337	5.96	5.89	5.95	5.90	0.11
Control questions second attempt (max=6)	5.99	337	6.00	5.97	5.99	5.98	0.22
All control questions correct	0.99	337	1.00	0.98	0.99	0.99	0.20
N	337	337	167	170	168	169	

Notes: Descriptive statistics of the majority sample (Column 1). Columns 3-6 present means for subjects in the specified experimental conditions. Experimental balance is tested in Column 7 using an F-test, showing that dummy variables "Wrongdoer OTHER" and "Scapegoat OTHER" do not significantly predict the specified descriptive variable. Variable "All control questions correct" indicates that the subject answered all control questions correctly at first or second attempt. For all variables, the values are missing for unspecified answers and for "I do not know" answers; Column 2 gives the number of non-missing values.

TABLE A2: DESCRIPTIVE STATISTICS AND RANDOMIZATION CHECKS (ROMA MINORITY SAMPLE)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Means across treatments				
Sample	Whole sample	N	Wrongdoer SAME	Wrongdoer OTHER	Scapegoat SAME	Scapegoat OTHER	Diff (F-test)
Female	0.55	484	0.54	0.57	0.56	0.55	0.76
Age	20.23	484	20.36	20.10	20.17	20.28	0.25
Student (secondary school)	0.14	484	0.12	0.16	0.16	0.12	0.12
Student (university)	0.01	484	0.02	0.01	0.02	0.01	0.48
Employed	0.17	484	0.18	0.16	0.18	0.16	0.66
Unemployed	0.49	484	0.49	0.48	0.46	0.51	0.42
At home	0.19	484	0.19	0.19	0.18	0.20	0.90
Primary school degree	0.57	484	0.56	0.58	0.56	0.58	0.79
Secondary school degree	0.43	484	0.44	0.42	0.44	0.42	0.79
Married	0.31	484	0.29	0.33	0.32	0.31	0.59
Child	0.48	484	0.50	0.45	0.48	0.47	0.61
Household size	6.02	484	6.00	6.05	6.09	5.95	0.88
Regularly goes to church	0.27	484	0.26	0.28	0.24	0.29	0.43
Father with secondary school	0.39	401	0.37	0.42	0.39	0.39	0.60
Father with university	0.04	401	0.03	0.05	0.04	0.04	0.47
Mother with secondary school	0.32	414	0.32	0.33	0.33	0.32	0.97
Mother with university	0.03	414	0.03	0.02	0.02	0.03	0.67
Father employed	0.55	392	0.55	0.55	0.56	0.53	0.87
Mother employed	0.38	436	0.41	0.35	0.39	0.38	0.37
Family owns a car	0.32	484	0.32	0.31	0.32	0.31	0.93
Family owns a computer	0.52	484	0.50	0.54	0.53	0.51	0.69
Family owns a smartphone	0.85	484	1.06	0.65	1.06	0.65	0.37
Family owns a tablet	0.39	484	0.36	0.42	0.42	0.37	0.18
Control questions first attempt (max=6)	4.27	483	4.18	4.36	4.16	4.38	0.19
Control questions second attempt (max=6)	5.53	482	5.47	5.58	5.53	5.52	0.50
All control questions correct	0.77	484	0.76	0.78	0.77	0.77	0.88
N	484	484	241	243	241	243	

Notes: Descriptive statistics of the Roma minority sample (Column 1). Columns 3-6 present means for subjects in the specified experimental conditions. Experimental balance is tested in Column 7 using an F-test, showing that dummy variables "Wrongdoer OTHER" and "Scapegoat OTHER" do not significantly predict the specified descriptive variable. Variable "All control questions correct" indicates that the subject answered all control questions correctly at first or second attempt. For all variables, the values are missing for unspecified answers and for "I do not know" answers; Column 2 gives the number of non-missing values.

TABLE A3: PUNISHMENT OF THE WRONGDOER, ROBUSTNESS CHECKS (MAJORITY SAMPLE)

Dependent variable	Punishment of the Wrongdoer (intensity)								
Sample	All	All	All	All	All	All	All	Subjects who answered all control questions correctly	Excluding potentially hypothetical observations
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Harm intensity	0.56***	0.56***	0.56***	0.56***	0.56***	0.56***	0.56***	0.56***	0.62***
	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)
Wrongdoer OTHER	-0.26**	-0.24**	-0.28**	-0.26**	-0.27**	-0.28**	-0.31**	-0.28**	-0.35***
	(0.11)	(0.11)	(0.12)	(0.11)	(0.12)	(0.11)	(0.13)	(0.11)	(0.13)
Harm intensity*Wrongdoer OTHER	0.14***	0.14***	0.14***	0.14***	0.14***	0.14***	0.14***	0.14***	0.14***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Controls									
Baseline controls	x		x	x	x	x	x	x	x
Design features			x				x		
Sample University (vs. High school)				x			x		
Location fixed effects					x		x		
Education of parents						x	x		
Mean baseline (Wrongdoer SAME, 0 harm)	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.37
Observations	1,685	1,685	1,685	1,685	1,685	1,685	1,685	1,670	1,241
R-squared	0.417	0.408	0.429	0.420	0.440	0.419	0.453	0.417	0.504

Notes: OLS, standard errors (in parentheses) are clustered at a Punisher level. *** denotes p<0.01, ** p<0.05, and * p<0.1. The dependent variable is the extent of punishment towards the Wrongdoer (EUR 0-8). "Harm intensity" captures the harm caused by the Wrongdoer to the Victim (EUR 0-8). Punishers (and Victims) are from the majority ethnic group. "Wrongdoer OTHER" indicates that the Wrongdoer comes from a different ethnic group (Roma minority) than the Punisher. Baseline controls include gender and age of the Punisher, and a dummy variable indicating that the Punishing the Wrongdoer Game took place before the Punishing the Scapegoat Game. Design features include a dummy variable "Scapegoat OTHER" (equal to one if the Scapegoat comes from a different ethnic group than the Punisher) and experimenter fixed effects. Education of parents includes dummy variables for mother/father with a university degree, and dummy variables for unknown education. In Column 8, we exclude all subjects who did not answer all control questions correctly at first or second attempt. In Column 9, we exclude all observations for harm levels, which a decision-maker believed were chosen by none of the twenty potential Wrongdoers.

TABLE A4: PUNISHMENT OF THE SCAPEGOAT, ROBUSTNESS CHECKS (MAJORITY SAMPLE)

Dependent variable	Punishment of the Scapegoat (intensity)								
Sample	All	All	All	All	All	All	All	Subjects who answered all control questions correctly	Excluding potentially hypothetical observations
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Harm intensity	0.08***	0.08***	0.08***	0.08***	0.08***	0.08***	0.08***	0.08***	0.11***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Scapegoat OTHER	-0.00	-0.01	-0.01	-0.00	-0.01	-0.01	-0.03	-0.02	0.03
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)	(0.08)	(0.09)
Harm intensity*Scapegoat OTHER	0.09***	0.09***	0.09***	0.09***	0.09***	0.09***	0.09***	0.08***	0.08**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Wrongdoer OTHER	-0.14	-0.14	-0.15	-0.14	-0.14	-0.12	-0.12	-0.16	-0.13
	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.15)
Controls									
Baseline controls	x		x	x	x	x	x	x	x
Design features			x				x		
Sample University (vs. High school)				x			x		
Location fixed effects					x		x		
Education of parents						x	x		
Mean baseline (Scapegoat SAME, 0 harm)	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
Observations	1,685	1,685	1,685	1,685	1,685	1,685	1,685	1,670	1,241
R-squared	0.072	0.068	0.100	0.072	0.091	0.085	0.126	0.070	0.092

Notes: OLS, standard errors (in parentheses) are clustered at a Punisher level. *** denotes p<0.01, ** p<0.05, and * p<0.1. The dependent variable is the extent of punishment towards the Scapegoat (EUR 0-8). "Harm intensity" captures the harm caused by the Wrongdoer to the Victim (EUR 0-8). Punishers (and Victims) are from the majority ethnic group. "Scapegoat OTHER" indicates that the Scapegoat comes from a different ethnic group (Roma minority) than the Punisher. "Wrongdoer OTHER" indicates that the Wrongdoer comes from a different ethnic group that the Punisher. Baseline controls include gender and age of the Punisher, and a dummy variable indicating that the Punishing the Wrongdoer Game took place before the Punishing the Scapegoat Game. Design features include experimenter fixed effects. Education of parents includes dummy variables for mother/father with a university degree, and dummy variables for unknown education. In Column 8, we exclude all subjects who did not answer all control questions correctly at first or second attempt. In Column 9, we exclude all observations for harm levels, which a decision-maker believed were chosen by none of the twenty potential Wrongdoers.

TABLE A5: PUNISHMENT OF THE WRONGDOER AND THE SCAPEGOAT, NON-LINEAR SPECIFICATIONS (MAJORITY SAMPLE AND ROMA MINORITY SAMPLE)

Dependent variable	Punishment of the Wrongdoer (intensity)	Punishment of the Scapegoat (intensity)	Punishment of the Wrongdoer (intensity)	Punishment of the Scapegoat (intensity)
Sample	Majority		Roma minority	
	(1)	(2)	(3)	(4)
Small harm (2 euro)	1.27***	0.30***	0.65***	0.41***
	(0.12)	(0.08)	(0.11)	(0.11)
Medium harm (4 euro)	2.35***	0.36***	1.24***	0.66***
	(0.17)	(0.09)	(0.15)	(0.13)
Large harm (6 euro)	3.72***	0.56***	1.39***	0.77***
	(0.23)	(0.12)	(0.18)	(0.15)
Maximum harm (8 euro)	4.41***	0.67***	1.54***	0.67***
	(0.29)	(0.15)	(0.20)	(0.16)
Wrongdoer OTHER	-0.24**	-0.14	-0.02	0.04
	(0.11)	(0.13)	(0.16)	(0.17)
Wrongdoer OTHER*Small harm (2 euro)	0.20		0.07	
	(0.15)		(0.17)	
Wrongdoer OTHER*Medium harm (4 euro)	0.63***		-0.01	
	(0.22)		(0.22)	
Wrongdoer OTHER*Large harm (6 euro)	0.72**		0.31	
	(0.30)		(0.26)	
Wrongdoer OTHER*Maximum harm (8 euro)	1.16***		0.50	
	(0.38)		(0.31)	
Scapegoat OTHER		-0.02		0.00
		(0.07)		(0.17)
Scapegoat OTHER*Small harm (2 euro)		0.13		0.02
		(0.11)		(0.16)
Scapegoat OTHER*Medium harm (4 euro)		0.42***		-0.01
		(0.15)		(0.18)
Scapegoat OTHER*Large harm (6 euro)		0.65***		-0.09
		(0.20)		(0.21)
Scapegoat OTHER*Maximum harm (8 euro)		0.62***		0.15
		(0.23)		(0.24)
Controls	baseline	baseline	baseline	baseline
Mean baseline	0.35	0.13	0.90	0.97
Observations	1,685	1,685	2,420	2,420
R-squared	0.419	0.074	0.076	0.019

Notes: OLS, standard errors (in parentheses) are clustered at a Punisher level. *** denotes p<0.01, ** p<0.05, and * p<0.1. The dependent variable in Columns 1 and 3 is the extent of punishment towards the Wrongdoer (EUR 0-8), the dependent variable in Columns 2 and 4 is the extent of punishment towards the Scapegoat (EUR 0-8). Punishers (and Victims) in Columns 1 and 2 are from the majority ethnic group, Punishers (and Victims) in Columns 3 and 4 are from the Roma ethnic group. "Small harm (2 euro)", "Medium harm (4 euro)", "Large harm (6 euro)", and "Maximum harm (8 euro) are dummy variables specifying the level of harm caused by the Wrongdoer to the Victim; zero harm (0 euro) is the omitted category. "Wrongdoer OTHER"/"Scapegoat OTHER" indicate that the Wrongdoer/the Scapegoat comes from a different ethnic group than the Punisher. Baseline controls include gender and age of the Punisher, and a dummy variable indicating that the Punishing the Wrongdoer Game took place before the Punishing the Scapegoat Game.

TABLE A6: PUNISHMENT OF THE WRONGDOER, HETEROGENEITY ANALYSIS (MAJORITY SAMPLE)

Dependent variable	Punishment of the Wrongdoer (intensity)					
Sample	Men	Women	High school sample	University sample	Parents with at most secondary education	At least one parent with at university education
	(1)	(2)	(3)	(4)	(5)	(6)
Harm intensity	0.61***	0.51***	0.62***	0.49***	0.57***	0.56***
	(0.05)	(0.05)	(0.04)	(0.06)	(0.04)	(0.06)
Wrongdoer OTHER	-0.18	-0.37***	-0.22*	-0.40*	-0.21	-0.35
	(0.18)	(0.14)	(0.13)	(0.21)	(0.13)	(0.21)
Harm intensity*Wrongdoer OTHER	0.14**	0.12*	0.07	0.23***	0.11*	0.18**
	(0.06)	(0.07)	(0.06)	(0.07)	(0.06)	(0.08)
Controls	baseline	baseline	baseline	baseline	baseline	baseline
Mean baseline (Wrongdoer SAME, 0 harm)	0.30	0.40	0.23	0.51	0.28	0.46
Observations	980	705	960	725	975	685
R-squared	0.477	0.355	0.451	0.404	0.403	0.446

Notes: OLS, standard errors (in parentheses) are clustered at a Punisher level. *** denotes p<0.01, ** p<0.05, and * p<0.1. The dependent variable is the extent of punishment towards the Wrongdoer (EUR 0-8). Columns 1 and 2 report results by the gender of the Punisher. Columns 3 and 4 report results for Punishers who are secondary school students and university students, respectively. Columns 5 and 6 report results for Punishers who have both parents with at most secondary school education and those who have at least one parent with a university degree, respectively (self-reported, omitting subjects with missing answers). "Harm intensity" captures the harm caused by the Wrongdoer to the Victim (EUR 0-8). Punishers (and Victims) are from the majority ethnic group. "Wrongdoer OTHER" indicates that the Wrongdoer comes from a different ethnic group (Roma minority) than the Punisher. Baseline controls include gender and age of the Punisher, and a dummy variable indicating that the Punishing the Wrongdoer Game took place before the Punishing the Scapegoat Game.

TABLE A7: PUNISHMENT OF THE SCAPEGOAT, HETEROGENEITY ANALYSIS (MAJORITY SAMPLE)

Dependent variable	Punishment of the Scapegoat (intensity)					
Sample	Men	Women	High school	University	Parents with at most secondary education	At least one parent with at university education
	(1)	(2)	(3)	(4)	(5)	(6)
Harm intensity	0.05**	0.11***	0.09***	0.06**	0.08***	0.07**
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
Scapegoat OTHER	-0.10	0.09	-0.04	-0.00	0.05	-0.09
	(0.12)	(0.10)	(0.09)	(0.15)	(0.08)	(0.16)
Harm intensity*Scapegoat OTHER	0.13***	0.03	0.08**	0.10**	0.09**	0.09*
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
Wrongdoer OTHER	-0.01	-0.35*	-0.15	-0.15	0.01	-0.32
	(0.19)	(0.19)	(0.15)	(0.22)	(0.17)	(0.22)
Controls	baseline	baseline	baseline	baseline	baseline	baseline
Mean baseline (Scapegoat SAME, 0 harm)	0.19	0.05	0.08	0.19	0.06	0.24
Observations	980	705	960	725	975	685
R-squared	0.074	0.094	0.101	0.084	0.092	0.063

Notes: OLS, standard errors (in parentheses) are clustered at a Punisher level. *** denotes p<0.01, ** p<0.05, and * p<0.1. The dependent variable is the extent of punishment towards the Scapegoat (EUR 0-8). Columns 1 and 2 report results by the gender of the Punisher. Columns 3 and 4 report results for Punishers who are secondary school students and university students, respectively. Columns 5 and 6 report results for Punishers who have both parents with at most secondary school education and those who have at least one parent with a university degree, respectively (self-reported, omitting subjects with missing answers). "Harm intensity" captures the harm caused by the Wrongdoer to the Victim (EUR 0-8). Punishers (and Victims) are from the majority ethnic group. "Scapegoat OTHER" indicates that the Scapegoat comes from a different ethnic group (Roma minority) than the Punisher. "Wrongdoer OTHER" indicates that the Wrongdoer comes from a different ethnic group that the Punisher. Baseline controls include gender and age of the Punisher, and a dummy variable indicating that the Punishing the Wrongdoer Game took place before the Punishing the Scapegoat Game.

TABLE A8: BELIEFS REGARDING THE BEHAVIOR OF THE WRONGDOER (MAJORITY SAMPLE)

	Harm	Wrongdoer SAME	Wrongdoer OTHER	Diff (Fisher exact test)
	(1)	(2)	(3)	(4)
Panel A: Beliefs Task 1—Wrongdoer destroys:	0	60.5	63.5	
	2	18.6	14.7	
	4	11.4	11.2	
	6	3.6	2.4	
	8	6.0	8.2	
				0.759
	Harm	Wrongdoer SAME	Wrongdoer OTHER	Diff (ranksum)
	(1)	(2)	(3)	(4)
Panel B: Beliefs Task2—Out of 20 Wrongdoers, how many destroy:	0	8.6	8.5	0.860
	2	3.4	3.0	0.322
	4	3.2	2.9	0.441
	6	2.2	2.4	0.403
	8	2.6	3.1	0.015

Notes: Beliefs of the Punishers from the majority ethnic group regarding which harm levels were chosen by the Wrongdoers. In Task 1 (Panel A), the Punishers were asked to guess which action was chosen by the Wrongdoer and rewarded 1 euro for a correct answer. The shares for majority ethnicity Wrongdoers and Roma ethnicity Wrongdoers are reported in Columns 2 and 3, respectively. The equality of the distributions is tested in Column 4 using a Fisher exact test. In Task 2 (Panel B), the Punishers were asked to guess how many of the twenty potential Wrongdoers chose each level of harm. The estimated number of Wrongdoers is reported in Columns 2 and 3 for the Wrongdoers of majority ethnicity and for the Wrongdoers of Roma ethnicity, respectively (we exclude 20 subjects whose reports do not sum up to twenty). Differences between conditions are tested using the Wilcoxon rank-sum test; p-values are specified in Column 4.

TABLE A9: PUNISHMENT OF THE WRONGDOER (ROMA MINORITY SAMPLE)

Dependent variable	Punishment of the Wrongdoer (intensity)	Punishment of the Wrongdoer (yes)	Punishment of the Wrongdoer (intensity)	Punishment of the Wrongdoer (intensity)	Punishment of the Wrongdoer (intensity)
Sample	All	All	Punishment of the Wrongdoer = yes	All	All CQ correctly
	(1)	(2)	(3)	(4)	(5)
Harm intensity	0.19***	0.02***	0.26***	0.19***	0.21***
	(0.03)	(0.00)	(0.04)	(0.03)	(0.03)
Wrongdoer OTHER	-0.09	-0.03	0.06	-0.15	-0.16
	(0.16)	(0.03)	(0.26)	(0.16)	(0.16)
Harm intensity*Wrongdoer OTHER	0.06*	0.01	0.06	0.06*	0.09**
	(0.04)	(0.01)	(0.05)	(0.04)	(0.04)
Controls	baseline	baseline	baseline	full	baseline
Mean baseline (Wrongdoer SAME, 0 harm)	0.90	0.27	3.41	0.90	0.79
Observations	2,420	2,420	1,141	2,420	1,860
R-squared	0.073	0.032	0.142	0.153	0.094

Notes: OLS, standard errors (in parentheses) are clustered at a Punisher level. *** denotes p<0.01, ** p<0.05, and * p<0.1. The dependent variable in Columns 1,3,4 is the extent of punishment towards the Wrongdoer (EUR 0-8). In Column 2, the dependent variable indicates that the Punisher chose non-zero punishment of the Wrongdoer. "Harm intensity" captures the harm caused by the Wrongdoer to the Victim (EUR 0-8). Punishers (and Victims) are from the Roma ethnic group. "Wrongdoer OTHER" indicates that the Wrongdoer comes from a different ethnic group (majority) than the Punisher. Baseline controls include gender and age of the Punisher, and a dummy variable indicating that the Punishing the Wrongdoer Game took place before the Punishing the Scapegoat Game. Full controls include also a dummy variable "Scapegoat OTHER" (equal to one if the Scapegoat comes from a different ethnic group than the Punisher), experimenter fixed effects, location fixed effects, education of parents (dummy variables for mother/father with at least secondary school education, dummy variables for unknown education), and a dummy variable indicating that the subject answered all control questions correctly at the first or second attempt. In Column 5, we exclude all subjects who did not answer all control questions correctly at the first or second attempt.

TABLE A10: PUNISHMENT OF THE SCAPEGOAT (ROMA MINORITY SAMPLE)

Dependent variable	Punishment of the Scapegoat (intensity)	Punishment of the Scapegoat (yes)	Punishment of the Scapegoat (intensity)	Punishment of the Scapegoat (intensity)	Punishment of the Scapegoat (intensity)
Sample	All	All	Punishment of the Scapegoat = yes	All	All CQ correctly
	(1)	(2)	(3)	(4)	(4)
Harm intensity	0.09***	0.01***	0.16***	0.09***	0.10***
	(0.02)	(0.00)	(0.04)	(0.02)	(0.02)
Scapegoat OTHER	-0.02	0.03	-0.38	0.01	-0.12
	(0.16)	(0.04)	(0.27)	(0.16)	(0.17)
Harm intensity*Scapegoat OTHER	0.01	-0.00	0.03	0.01	0.01
	(0.03)	(0.00)	(0.05)	(0.03)	(0.03)
Wrongdoer OTHER	0.04	-0.02	0.27	-0.01	0.04
	(0.17)	(0.04)	(0.18)	(0.16)	(0.18)
Controls	baseline	baseline	baseline	full	baseline
Mean baseline (Scapegoat SAME, 0 harm)	0.97	0.26	3.77	0.97	0.80
Observations	2,420	2,420	906	2,420	1,860
R-squared	0.017	0.008	0.064	0.093	0.021

Notes: OLS, standard errors (in parentheses) are clustered at a Punisher level. *** denotes p<0.01, ** p<0.05, and * p<0.1. The dependent variable in Columns 1,3,4 is the extent of punishment towards the Scapegoat (EUR 0-8). In Column 2, the dependent variable indicates whether the Punisher chose non-zero punishment of the Scapegoat. "Harm intensity" captures the harm caused by the Wrongdoer to the Victim (EUR 0-8). Punishers (and Victims) are from the Roma ethnic group. "Scapegoat OTHER" indicates that the Scapegoat comes from a different ethnic group (majority) than the Punisher. "Wrongdoer OTHER" indicates that the Wrongdoer comes from a different ethnic group that the Punisher. Baseline and full controls are defined in Table A9. In Column 5, we exclude all subjects who did not answer all control questions correctly at the first or second attempt.

TABLE A11: PERCEIVED SOCIO-ECONOMIC STATUS OF SAME/OTHER WRONGDOERS AND SCAPEGOATS (MAJORITY SAMPLE)

Dependent variable	Wrongdoer Unemployed (perceived)	Wrongdoer Student (perceived)	Wrongdoer's father low education (perceived)	Scapegoat Unemployed (perceived)	Scapegoat Student (perceived)	Scapegoat's father low education (perceived)
Sample	All	All	All	All	All	All
	(1)	(2)	(3)	(4)	(5)	(6)
Wrongdoer OTHER	0.52***	-0.69***	0.65***			
	(0.04)	(0.04)	(0.04)			
Scapegoat OTHER				0.40***	-0.61***	0.63***
				(0.04)	(0.04)	(0.04)
Constant	0.07***	0.85***	0.02**	0.08***	0.78***	0.02*
	(0.02)	(0.03)	(0.01)	(0.02)	(0.03)	(0.01)
Controls	no	no	no	no	no	no
Observations	337	337	337	337	337	337
R-squared	0.309	0.478	0.466	0.201	0.371	0.443

Notes: OLS, robust standard errors in parentheses. *** denotes p<0.01, ** p<0.05, and * p<0.1. The dependent variables are perceived socio-economic characteristics of the Wrongdoer (Column 1-3) and the Scapegoat (Column 4-6), elicited from the Punishers at the end of the experiment. Punishers are from the majority ethnic group. "Wrongdoer OTHER"/"Scapegoat OTHER" indicates that the Wrongdoer/the Scapegoat come from a different ethnic group (Roma minority) than the Punisher.

TABLE A12: PUNISHMENT OF THE WRONGDOER, CONTROLLING FOR THE PERCEIVED SOCIO-ECONOMIC ATTRIBUTES (MAJORITY SAMPLE)

Dependent variable	Punishment of the Wrongdoer (intensity)			
Sample	All	All	All	All
	(1)	(2)	(3)	(4)
Harm intensity	0.56***	0.56***	0.56***	0.56***
	(0.03)	(0.03)	(0.03)	(0.03)
Wrongdoer OTHER	-0.26**	-0.33**	-0.23*	-0.29*
	(0.11)	(0.16)	(0.14)	(0.17)
Harm intensity*Wrongdoer OTHER	0.14***	0.11*	0.14**	0.11
	(0.05)	(0.06)	(0.06)	(0.07)
Wrongdoer Unemployed (perceived)		0.13		0.14
		(0.17)		(0.18)
Harm intensity*Wrongdoer Unemployed		0.07		0.07
		(0.06)		(0.06)
Wrongdoer's father low education (perceived)			-0.04	-0.07
			(0.13)	(0.13)
Harm intensity*Wrongdoer's father low education			0.01	-0.01
			(0.06)	(0.06)
Controls	baseline	baseline	baseline	baseline
Mean baseline (Wrongdoer SAME, 0 harm)	0.35	0.35	0.35	0.35
Observations	1,685	1,685	1,685	1,685
R-squared	0.417	0.420	0.417	0.420

Notes: OLS, standard errors (in parentheses) are clustered at a Punisher level. *** denotes p<0.01, ** p<0.05, and * p<0.1. The dependent variable is the extent of punishment towards the Wrongdoer (EUR 0-8). "Harm intensity" captures the harm caused by the Wrongdoer to the Victim (EUR 0-8). Punishers (and Victims) are from the majority ethnic group. "Wrongdoer OTHER" indicates that the Wrongdoer comes from a different ethnic group (Roma minority) than the Punisher. "Wrongdoer Unemployed (perceived)" indicates that the Punisher believes that the (SAME or OTHER) Wrongdoer is unemployed. "Wrongdoer's father low education (perceived)" indicates that the Punisher believes that the (SAME or OTHER) Wrongdoer has a father with at most primary school education. Perceived socio-economic characteristics of the Wrongdoer were elicited from the Punishers at the end of the experiment. Baseline controls include gender and age of the Punisher, and a dummy variable indicating that the Punishing the Wrongdoer Game took place before the Punishing the Scapegoat Game.

TABLE A13: PUNISHMENT OF THE SCAPEGOAT, CONTROLLING FOR THE PERCEIVED SOCIO-ECONOMIC ATTRIBUTES (MAJORITY SAMPLE)

Dependent variable	Punishment of the Scapegoat (intensity)			
Sample	All	All	All	All
	(1)	(2)	(3)	(4)
Harm intensity	0.08***	0.08***	0.08***	0.08***
	(0.02)	(0.02)	(0.02)	(0.02)
Scapegoat OTHER	-0.00	-0.08	-0.04	-0.08
	(0.08)	(0.08)	(0.10)	(0.10)
Harm intensity*Scapegoat OTHER	0.09***	0.07**	0.08**	0.07*
	(0.03)	(0.03)	(0.04)	(0.04)
Scapegoat Unemployed (perceived)		0.18**		0.18**
		(0.09)		(0.09)
Harm intensity*Scapegoat Unemployed		0.04		0.04
		(0.04)		(0.04)
Scapegoat's father low education (perceived)			0.06	0.01
			(0.09)	(0.09)
Harm intensity*Scapegoat's father low education			0.01	0.00
			(0.04)	(0.04)
Wrongdoer OTHER	-0.14	-0.10	-0.13	-0.10
	(0.13)	(0.13)	(0.13)	(0.13)
Controls	baseline	baseline	baseline	baseline
Mean baseline (Scapegoat SAME, 0 harm)	0.13	0.13	0.13	0.13
Observations	1,685	1,685	1,685	1,685
R-squared	0.072	0.080	0.073	0.080

Notes: OLS, standard errors (in parentheses) are clustered at a Punisher level. *** denotes p<0.01, ** p<0.05, and * p<0.1. The dependent variable is the extent of punishment towards the Scapegoat (EUR 0-8). "Harm intensity" captures the harm caused by the Wrongdoer to the Victim (EUR 0-8). Punishers (and Victims) are from the majority ethnic group. "Scapegoat OTHER" indicates that the Scapegoat comes from a different ethnic group (Roma minority) than the Punisher. "Scapegoat Unemployed (perceived)" indicates that the Punisher believes that the (SAME or OTHER) Scapegoat is unemployed. "Scapegoat's father low education (perceived)" indicates that the Punisher believes that the (SAME or OTHER) Scapegoat has a father with at most primary school education. "Wrongdoer OTHER" indicates that the Wrongdoer comes from a different ethnic group that the Punisher. Perceived socio-economic characteristics of the Scapegoat were elicited from the Punishers at the end of the experiment. Baseline controls include gender and age of the Punisher, and a dummy variable indicating that the Punishing the Wrongdoer Game took place before the Punishing the Scapegoat Game.

[FOR ONLINE PUBLICATION]

Scapegoating: Experimental Evidence

Michal Bauer, Jana Cahlíková, Julie Chytilová, Gérard Roland, and Tomáš Želinský

Supplementary Online Appendix B

This file contains the experimental protocol

EXPERIMENTAL INSTRUCTIONS – PUNISHER

[English translation; instructions were given in Slovak language, explained to the subjects one-on-one]

[The order of the Punishing the Wrongdoer Game and Punishing the Scapegoat Game was randomized. The baseline version of the instructions below are for the case when Punishing the Wrongdoer Game was implemented first, text in blue color shows the adjustment for when the Punishing the Scapegoat Game was implemented first.]

1. Introduction

[After subjects come, check their IDs and make sure they are 18 or older.]

Good morning, please take a seat. My name is XXX. Today you will go through several tasks that will together take about 45 minutes. You will get at least 8 euro and 20 cents for completing the tasks. The money will be paid to you after we finish. If you decide that you don't want to continue with the tasks, you can leave at any time. If you leave before the end, you will get 2 euro. In the tasks, you will be making decisions. Your decisions will stay anonymous; we will never tell anybody how you decided. Now please read this and sign that you are willing to participate in the tasks.

[Send subjects to an experimenter]

[Before the subject comes, fill the introductory screen of the app according to the Log file]

[When the subject comes, the app is at the "START" screen]

Please, take a seat. My name is XX and we'll go through several tasks together. At the beginning, you get 9 euro from us. This amount can be slightly reduced, which depends solely on your decisions. You will get at least 8 euro and 20 cents.

Please listen to the instructions carefully. I want you to understand everything before we begin. You can also ask questions at any time.

A few weeks ago we recruited people from several locations in Eastern Slovakia to work for us.

Today, I will show you photos of these people from three locations.

*[Move to the next screen – with photos **/1: Introduction/**]*

For simplicity, I will call these photos A, B and C. When I talk about a Person A, I have in mind a Roma/non-Roma person from Photo A *[point to Photo A]*, but we do not know exactly which person from that photo. Similarly, when I talk about Person B, I have in mind a Roma/non-Roma person from Photo B *[point to Photo B]*. And finally, when I talk about Person C I have in mind a Roma/non-Roma person from Photo C *[point to Photo C]*. Since each photo was taken at a different location, Person A, B and C do not know each other.

Everybody completed the work and earned 8 euro. Everybody also knew her/his earnings may depend on choices of other people.

*[Move to the next screen – with photos **/2: Introduction/**]*

After they finished working, Person A saw the Photo B [*point to the photos*]. Person A was told that Person B also worked and earned 8 euro for the same job, but worked at a different location. Person A did not get any other information about Person B. Then, Person A got a chance to harm Person B. Person A could choose how much of the money earned by Person B he wanted to destroy.

*[Move to the next screen – with photos /**3: Introduction**/ - stay on this screen until control questions]*

Person A could choose one out of five possibilities: destroy all 8 euro of Person B, destroy 6 euro, destroy 4 euro, destroy 2 euro or not destroy anything and keep Person B's money unchanged at 8 euro. Person B did not see anybody's photo and could not harm anybody. Put differently, Person A's money stayed at 8 euro, whereas Person B gets 0, 2, 4, 6 or 8 euro—depending on whether and by how much the Person A decided to harm him.

In addition, there is also Person C from yet a different location *[point to Photo C]*. Person C also earned 8 euro for working for us. Person C did not see anybody's photo, could not harm anybody and could be harmed neither by Person A nor by Person B.

To summarize: All three persons – A, B, and C worked and earned 8 euro for that. Person A was the only person who saw a photo of somebody else – he saw Person B's photo and could harm Person B by destroying part or all of his income. Persons B and C didn't see anybody's photo, and at the same time, could not harm anybody.

Do you have any questions?

Now I will ask you some questions to make sure that you understand and remember the situation I just described. This summary of the situation can help you answer the questions *[point to the screen]*.

[Record the answers and if the participant did not fully understand, describe the setting again and ask the comprehension questions one more time, record the answers.]

1. Do Persons A, B, and C come from the same or different locations in Eastern Slovakia?
2. Do Persons A, B and C know each other?
3. How much money did Person A earn for the work? And Person B? And Person C?
4. Please name all persons who saw a photo of somebody else.
5. Please name all persons who could harm somebody else.
6. Please name all persons who could be harmed by somebody else.

*[Move to the next screen – **First task, and immediately tap "NEXT"**]*

2. *Task 1 – Punishing the Wrongoer Game/Punishing the Scapegoat Game*

[Note: Examples of decision screens are available in Section 7]

Now we will start with your Task.

Persons A, B, and C did not get the money yet, but we will pay them soon, after you make your decisions.

In this task, you can punish a Roma/non-Roma from Photo A for harming a Roma/non-Roma from Photo B. [In this task, you cannot punish a Roma/non-Roma from Photo A for harming a Roma/non-Roma from Photo B, but you can reduce the money of a Roma/non-Roma from Photo C.] You can choose one out of five options by tapping the "MINUS" or "PLUS" buttons, whereas your earnings are always displayed in the bottom right corner:
*[Illustrate on the tablet **by tapping the "Minus/Plus" buttons – tap to 0 first, then one by one explain all options. Draw subject's attention to their earnings displayed in the bottom right corner; it is not necessary to say the amounts out loud.**]*

- You can reduce the Person A's/Person C's money by 8 euro by paying 80 cents. Then Person A/Person C gets zero euros.
- You can reduce the Person A's/Person C's money by 6 euro by paying 60 cents. Then Person A/Person C gets 2 euro.
- You can reduce the Person A's/Person C's money by 4 euro by paying 40 cents. Then Person A/Person C gets 4 euro.
- You can reduce the Person A's/Person C's money by 2 euro by paying 20 cents. Then Person A/Person C gets 2 euro.
- Of course, you can also choose to keep the money of Person A/Person C untouched at 8 euro, for which you do not pay anything

You cannot change anybody else's money. Also, remember that nobody can reduce your payment. You will get between 8.20 and 9 euro for participating today, depending on your decisions. You will get the money immediately after we finish.

Person A has already decided whether and how much he wants to harm Person B, but for now, you don't know what he chose. I will ask you to make a decision for all five possible situations – for when Person A decided not to destroy any money and keep Person B's money unchanged at 8 euro, and for when he chose to harm Person B by destroying 2 euro, 4 euro, 6 euro, or all of his 8 euro.

Your choice in this task may have real consequences for Person A/Person C. In the end, we will randomly draw one out of ten balls from this bag ***[Show the bag. There are 10 balls in it: 8 blank ones, one with a star and one with a triangle]***. One of the ten balls has a star on it. If we draw the ball with the star, Person A/Person C will get the money according to your choice in this task. We will check what Person A chose and then see how you decided for that situation to calculate the final payments for Person A/Person C and for you. Since your choice in this task may have real consequences for Person A/Person C, please think carefully about your decision.

Your decision is completely anonymous. Person A/Person C will receive the money based on your decision, but will not receive any further information.

Do you have any questions?

First, we will go through a few examples:

*[Tap "NEXT": **Task 1 – Examples** + Next]*

*[Current screen: **Task 1 – Example 1** with pictures]*

1. A first example. **[*Let the subjects tap the screen themselves*].** In this situation, Person A chose to harm the Person B by destroying 4 euro, as you can see here in the picture. Therefore, at the moment, Person A's has 8 euro, Person B's money has 4 euro and Person C's has 8 euro *[show on the tablet]*. Now, you can decide whether and by how much you want to reduce the money of Person A/Person C. You cannot change anybody else's money. You can make the decision by simply tapping the "minus" and "plus" buttons to choose the preferred amount of money for Person A/Person C. Let's assume you decided to keep Person's A/Person C's money at 8 euro. In such a case Person A gets 8 euro, person B gets 4 euro and Person C gets 8 euro [*point at the earnings displayed on the screen*]. You get 9 euro [*point at the earnings displayed on the screen*]. Once you have decided, you can tap "OK" [*tap "OK" button now*], and a confirmation window with the chosen amount appears. If you agree, tap "OK" again. If you want to change your decision, tap "BACK".
   [Tap "OK" – move to example 2]
2. Another example: We are still in the situation in which Person A decided to destroy 4 euro of Peron B, as you can see here. Now assume that you decided to reduce Person's A/Person C's money by 8 euro. ***[Change Person's A/Person C's money to 0 euro.]*** Now, Person A gets 0 euro/8 euro, Person B gets 4 euro and person C gets 8 euro/0 euro [*point at the earnings displayed on the screen*]. Because you have to pay 80 cents to reduce Person's A/Person C' money, you get 8 euro and 20 cents [*point at the earnings displayed on the screen*]. Once you have decided, tap "OK" [*tap "OK" button now*], and a confirmation window with the chosen amount appears. If you agree, tap "OK" again. If you want to change your decision, tap "BACK".
   [Tap "OK" – move to example 3]
3. A different example: We are still in the situation in which Person A decided to destroy 4 euro of Peron B, as you can see here. Now assume that you decided to reduce Person's A/Person C's money by 4 euro. ***[Change Person's A/Person C's money to 4 euro.]*** In that case, Person A gets 4 euro/8 euro, Person B gets 4 euro and person C gets 8 euro/4 euro [*point at rewards on the screen*]. Because you have to pay 40 cents to reduce Person's A / Person's C money, you get 8 euro and 60 cents [*point to the earnings on the screen*]. Once you have decided, tap "OK" [*tap "OK" button now*] and confirm again.

Do you have any questions?

Now it's time for you to make the decisions. We will go, one by one, through all the five possible situations.

*[Move to the next screen – **Situation 1 – Explanation + Next**]*

*[Current screen: **Situation 1 – Explanation** with pictures]*

This is the first situation. In this situation, Person A decided not to destroy any of Person's B money and kept his money unchanged at 8 euro, as you can see here. At this moment, Person B thus has 8 euro *[point to the screen and **emphasize**!]*.
Now you can decide whether and by how much you want to reduce the money of Person A/Person C. You cannot change anybody else's money. Tap the "minus" and "plus" buttons as you need. When you are done, tap the "OK" button. **Beware all the situations before were examples, now this is your real decision which will count**. I will turn away when you decide, so that you can make your decision in complete privacy. Take your time with the decision, we have enough time. When you are done with the first situation, let me know, and we'll together proceed to a new situation.

*[Click OK. An introductory screen appears: **Situation 1 – Decision**]*

[Now hand the tablet to the participant, s/he clicks "Next", a screen with pictures appears, and s/he makes the decision.]

*[Current screen: **Situation 2 – Explanation** with pictures]*

This is the second situation. In this situation, Person A decided to destroy 2 euro of Person B, as you can see here. At this moment, Person B thus has 6 euro *[point to tablet and **emphasize**!]*.
Now you should decide whether and by how much you want to reduce the money of Person A/Person C. [Click *OK. An introductory screen appears: **Situation 2 – Decision**]*

[Now hand the tablet to the participant, s/he clicks "Next", a screen with pictures appears, and s/he makes the decision.]

*[Current screen: **Situation 3 – Explanation** with pictures]*

This is the third situation. In this situation, Person A decided to destroy 4 euro of Person B, as you can see here. At this moment, Person B thus has 4 euro *[point to tablet and **emphasize**!]*.
Now you should decide whether and by how much you want to reduce the money of Person A/Person C. [Click *OK. An introductory screen appears: **Situation 3– Decision**]*

[Now hand the tablet to the participant, s/he clicks "Next", a screen with pictures appears, and s/he makes the decision.]

*[Current screen: **Situation 4 – Explanation** with pictures]*

This is the fourth situation. In this situation, Person A decided to destroy 6 euro of Person B, as you can see here. At this moment, Person B thus has 2 euro *[point to tablet and **emphasize**!]*.
Now you should decide whether and by how much you want to reduce the money of Person A/Person C. [Click *OK. An introductory screen appears: **Situation 4– Decision**]*

[Now hand the tablet to the participant, s/he clicks "Next", a screen with pictures appears, and s/he makes the decision.]

*[Current screen: **Situation 5 – Explanation** with pictures]*

This is the fifth situation. In this situation, Person A decided to destroy all 8 euro of Person B, as you can see here. At this moment, Person B thus has no money *[point to tablet and **emphasize**!]*.
Now you should decide whether and by how much you want to reduce the money of Person A/Person C. [Click *OK. An introductory screen appears: **Situation 5– Decision**]*

[Now hand the tablet to the participant, s/he clicks "Next", a screen with pictures appears, and s/he makes the decision.]

*[Now move to the next screen – **Task 2 and "Next"**]*

3. Task 2 – Punishing the Scapegoat Game/Punishing the Wrongdoer Game

[Note: Examples of decision screens are available in Section 7]

Now we will move to a second task.

In this task, you cannot punish a Roma/non-Roma from Photo A for harming a Roma/non-Roma from Photo B, but you can reduce the money of a Roma/non-Roma from Photo C. [In this task, you can punish a Roma/non-Roma from Photo A for harming a Roma/non-Roma from Photo B.] Again, you can choose one out of five possibilities *[illustrate on the tablet **by tapping the "Minus/Plus" buttons – tap to 0 first, then one by one explain all options. Draw subject's attention to their earnings displayed in the bottom right corner**]:*

- You can reduce the Person C's/Person A's money by 8 euro by paying 80 cents. Then Person C/Person A gets zero euros.
- You can reduce the Person C's/Person A's money by 6 euro by paying 60 cents. Then Person C/Person A gets 2 euro.
- You can reduce the Person C's/Person A's money by 4 euro by paying 40 cents. Then Person C/Person A gets 4 euro.
- You can reduce the Person C's/Person A's money by 2 euro by paying 20 cents. Then Person C/Person A gets 2 euro.
- Of course, you can also choose to keep the money of Person C/Person A untouched at 8 euro, for which you do not pay anything

You cannot change anybody else's money. Also, remember that nobody can reduce your payment.

As before, I will ask you to make a decision for all five possible situations.

Your choice in this task may have real consequences for Person C/Person A. As I already told you, we will randomly draw one out of ten balls from this bag at the end ***[Show the bag. There are 10 balls in it: 8 blank ones, one with a star and one with a triangle]***. One of the ten balls has a triangle on it. If we draw the ball which has a triangle on it, Person C/Person A will get the money according to your choice in this task. Since your choice in this task may have real consequences for Person C/Person A, please think carefully about your decision.

Remember that only one of your decisions can have real consequences – either your decision about the money of the Person A/Person C – if we draw a ball with a star –or your decision about the money of the Person C/Person A – if we draw a ball with a triangle.

Your decision is completely anonymous. Person C/Person A will receive the money based on your decision, but will not receive any further information.

Do you have any questions?

Now, we will again go through a couple of examples:

*[Tap "NEXT": **Task 2 – Examples** + Next]*

*[Current screen: **Task 2 – Example 1** with pictures]*

In this situation, Person A decided to destroy 4 euro of Person B, as you can see here. Therefore, at the moment, Person A has 8 euro, Person B has 4 euro and Person C has 8 euro *[show on the tablet]*. Now, you can decide whether and by how much you want to reduce the money of Person C/Person A. You cannot change anybody else's money. You can again make the decision by simply tapping the "minus" and "plus" buttons.

[Go through all the examples, staying at the same screen, experimenter taps the buttons]

1. First example: First assume you decided to keep Person C's/Person A's money at 8 euro. In such a case Person A gets 8 euro, person B gets 4 euro and Person C gets 8 euro [*point at earnings displayed on the screen*]. You get 9 euro [*point at the earnings displayed on the screen*].
2. Another example: Now assume that you decided to reduce Person C's/Person A's money by 8 euro. ***[Change Person's C/Person A's money to 0 euro.]*** Now, Person A gets 8 euro/0 euro, Person B gets 4 euro and person C gets 0 euro/8 euro [*point at rewards on the screen*]. Because you have to pay 80 cents to reduce Person C's/Person A's money, you get 8 euro and 20 cents [*point at the earnings on the screen*].
3. A different example: Now assume that you decided to reduce Person C's/Person A's money by 4 euro. ***[Change Person's C/Person A's money to 4 euro.]*** Now, Person A gets 8 euro/4 euro, Person B gets 4 euro and person C gets 4 euro/8 euro [*point at rewards on the screen*]. Because you have to pay 40 cents to reduce Person C's/Person A's money, you get 8 euro and 60 cents [*point at the earnings on the screen*]. As before, once you have decided, you can tap "OK" [*tap "OK" button now*], a confirmation window with the chosen amount will appear. If you agree, tap "OK" again. If you want to change your decision, tap "BACK".

Do you have any questions?

Again, we will go, one by one, through all the five possible situations.

[Go through all the five situations]

*[Move to the next screen – **Situation 1 – Explanation + Next**]*

*[Current screen: **Situation 1 – Explanation** with pictures]*

This is the first situation. In this situation, Person A decided not to destroy any of Person's B money and kept his money unchanged at 8 euro, as you can see here. At this moment, Person B thus has 8 euro *[point to the screen and **emphasize**!]*.
Now you can decide whether and by how much you want to reduce the money of Person C/Person A. You cannot change anybody else's money.
Beware all the situations before were examples, now this is your real decision which will count. I will turn away when you decide, so that you can make your decision in complete privacy. Take your time with the decision, we have enough time. When you are done with the first situation, let me know, and we'll together proceed to a new situation.

*[Click OK. An introductory screen appears: **Situation 1 – Decision]***

[Now hand the tablet to the participant, s/he clicks "Next", a screen with pictures appears, and s/he makes the decision.]

*[Current screen: **Situation 2 – Explanation** with pictures]*

This is the second situation. In this situation, Person A decided to destroy 2 euro of Person B, as you can see here. At this moment, Person B thus has 6 euro *[point to tablet and **emphasize**!]*.
Now you should decide whether and by how much you want to reduce the money of Person C/Person A. [Click *OK. An introductory screen appears: **Situation 2 – Decision**]*

[Now hand the tablet to the participant, s/he clicks "Next", a screen with pictures appears, and s/he makes the decision.]

*[Current screen: **Situation 3 – Explanation** with pictures]*

This is the third situation. In this situation, Person A decided to destroy 4 euro of Person B, as you can see here. At this moment, Person B thus has 4 euro *[point to tablet and **emphasize**!]*.
Now you should decide whether and by how much you want to reduce the money of Person C/Person A. [Click *OK. An introductory screen appears: **Situation 3– Decision**]*

[Now hand the tablet to the participant, s/he clicks "Next", a screen with pictures appears, and s/he makes the decision.]

*[Current screen: **Situation 4 – Explanation** with pictures]*

This is the fourth situation. In this situation, Person A decided to destroy 6 euro of Person B, as you can see here. At this moment, Person B thus has 2 euro *[point to tablet and **emphasize**!]*.
Now you should decide whether and by how much you want to reduce the money of Person C/Person A. [Click *OK. An introductory screen appears: **Situation 4– Decision**]*

[Now hand the tablet to the participant, s/he clicks "Next", a screen with pictures appears, and s/he makes the decision.]

*[Current screen: **Situation 5 – Explanation** with pictures]*

This is the fifth situation. In this situation, Person A decided to destroy all 8 euro of Person B, as you can see here. At this moment, Person B thus has no money *[point to tablet and **emphasize**!]*.
Now you should decide whether and by how much you want to reduce the money of Person C/Person A. [Click *OK. An introductory screen appears: **Situation 5– Decision**]*

[***Now hand the tablet to the participant, s/he clicks "Next", a screen with pictures appears, and s/he makes the decision.]***

*[Now move to the next screen – **Task 3 and "Next"**]*

4. Task 3 – Beliefs elicitation

Now we will move to Task 3.

As I explained before, Person A could harm Person B. Person A could choose one out of five possibilities: destroy all 8 euro of Person B, destroy 6 euro, destroy 4 euro, destroy 2 euro or not destroy any money and keep Person B's money unchanged at 8 euro.

In this task, I will ask you to guess how Person A decided. If you guess correctly, we will add another euro to your earnings. You can see all the options on the tablet here. Tap on the option that you want to choose and then tap the "OK" button.

[Tap "OK", a new screen appears: Task 3 – Decision 1, hand the tablet to the participant, s/he clicks "Next" and makes a decision.]

[A new screen: **Task 3 – Decision 2**]

[Fill it in together with the participants.]

Now I will ask you a similar question, but in a greater detail. How many people out of those 20 in photo A do you think chose any of the following options? Tell me a number of people for each option and I'll write it down; overall, your responses should sum up to 20.

So, how many out of the 20 people do you think wanted to destroy all 8 euro of Person B? _____ How many wanted to destroy 6 euro? _____ How many wanted to destroy 4 euro? _____ How many wanted to destroy 2 euro? _____ And how many didn't want to destroy any money of Person B and wanted to keep his money unchanged at 8 euro? _____

[Fill in together with participants; it's not necessary for their responses to sum up to 20.]

5. Questionnaires

[Go through the screens together with the participants.]

[Task 1, Situation 5 – Assessment]

In this situation, where Person A decided to destroy all 8 euro of Person B, you decided to destroy X euro of Person A/Person C *[point at the screen]*. Why did you make such a decision?

[If the participant does not come up with a reason in 10 seconds:] Does anything come to your mind? If not, we can leave it blank.

[Task 1, Situation 5 – Fairness]

Imagine a situation, in which Person A decided to destroy all 8 euro of Person B. Somebody like you saw it and decided to destroy all 8 euro of Person A/Person C. [*Show the situation on the screen*] Do you find it very unfair, quite unfair, somewhat unfair, or somewhat fair, quite fair, or very fair?

[Task 2, Situation 5 – Assessment]

In this situation, where Person A decided to destroy all 8 euro of Person B, you decided to destroy X euro of Person C/Person A [*point at the screen]*. Why did you make such a decision?

[If the participant does not come up with a reason in 10 seconds:] Does anything come to your mind? If not, we can leave it blank.

[Task 2, Situation 5 – Fairness]

Imagine a situation, in which Person A decided to destroy all 8 euro of Person B. Somebody like you saw it and decided to destroy all 8 euro of Person C/Person A. [*Show the situation on the screen*] Do you find it very unfair, quite unfair, somewhat unfair, or somewhat fair, quite fair, or very fair?

[Person A, C – guesses]

Try to guess: What do you think, is Person A is a student, employed or unemployed?

Try to guess: What education does the father of Person A have?

Do you know anybody from this photograph A? **[If yes:]** How many people? From where?

Try to guess: What do you think, is Person C is a student, employed or unemployed?

Try to guess: What education does the father of Person C have?

Do you know anybody from this photograph C? **[If yes:]** How many people? From where?

[Demographic questionnaire]

Gender, age, education level, economic activity, marital status, number of children, going to church, education and employment status of the parents, number of people living in the same household, household possessions (car, personal computer/laptop, smartphone, tablet)

6. Conclusion

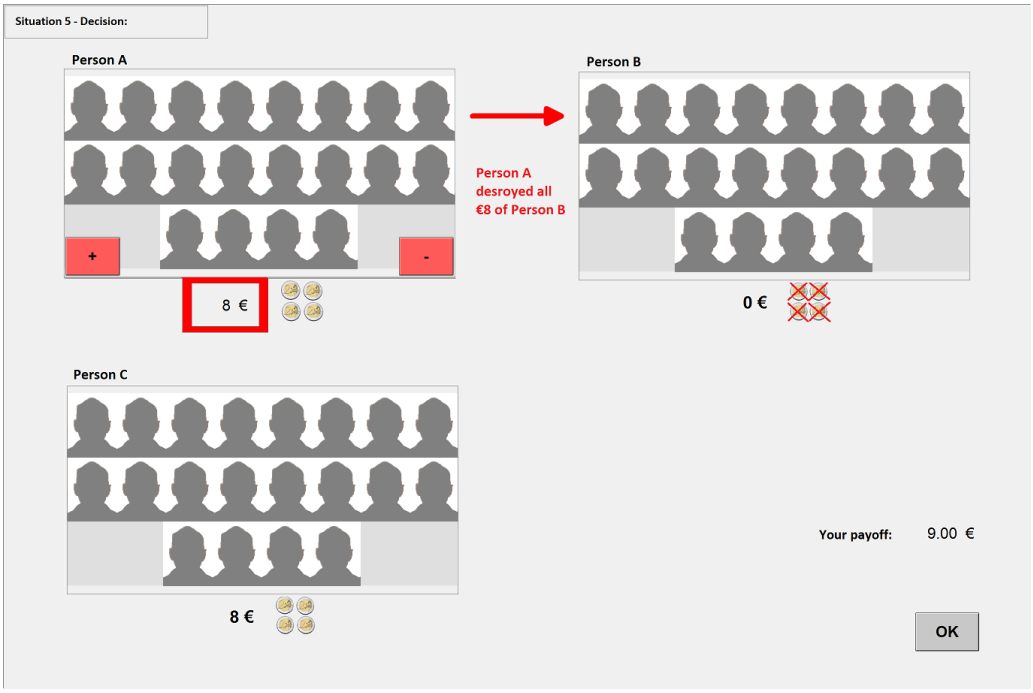
[Draw a ball to determine which task is payoff relevant, enter it into the app.]

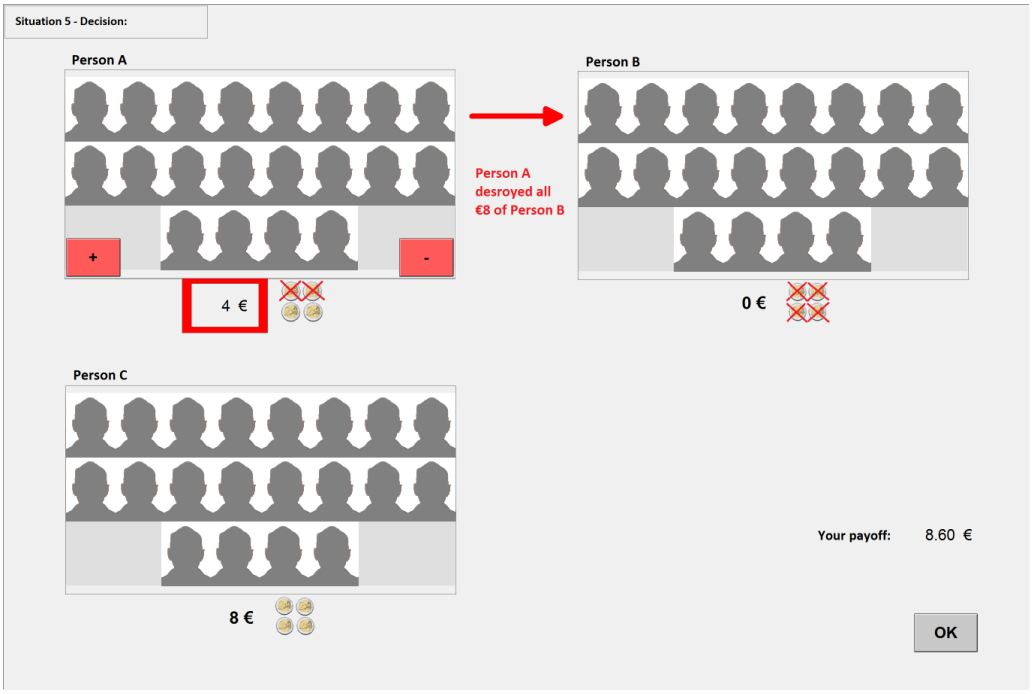
[Final earnings are shown on the screen, fill in the payoff receipt.]

That's it, thanks a lot for participating! Please go outside to my colleague who will give you your payment.

7. Examples of decision screens

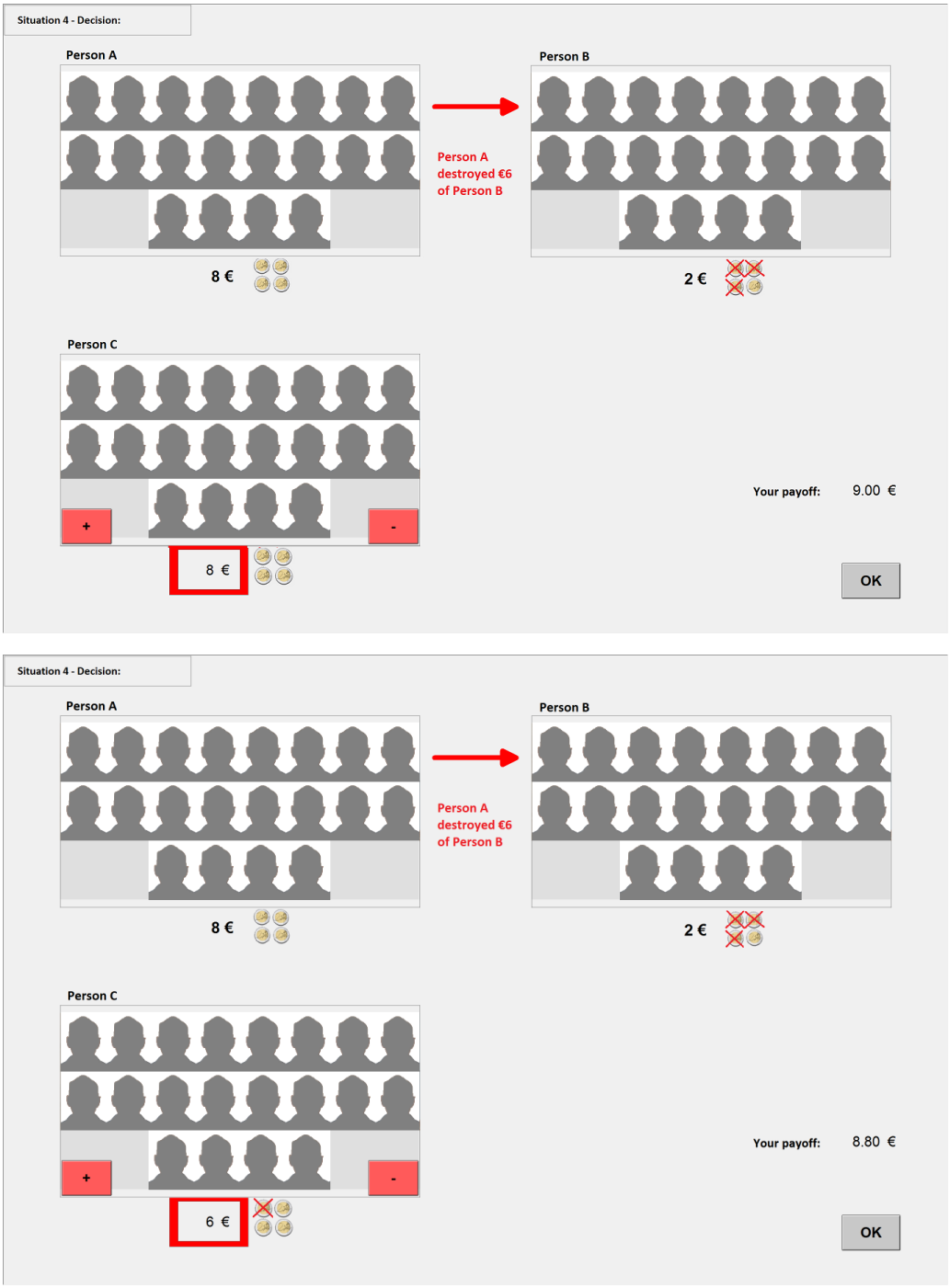
7.1. Punishing the Wrongdoer Game





Note: Examples of the decision screens for the Punishing the Wrongdoer Game, for the situation where the Wrongdoer (Person A) destroyed all €8 of the Victim (Person B). The default/starting point in all situations is always Person A and Person C getting all €8, as shown in the top figure. In the bottom figure, the Punisher is choosing to reduce €4 of the Wrongdoer and paying €0.40 for that (see the Punisher's earnings in the bottom right). In the experiment, each picture shows 20 different passport-style photographs, which vary based on the condition. For privacy reasons, we are not allowed to display the actual pictures of all study participants in publicly available materials.

7.2. Punishing the Scapegoat Game



Note: Examples of the decision screens for the Punishing the Scapegoat Game, for the situation where the Wrongdoer (Person A) destroyed €6 of the Victim (Person B). The default/starting point in all situations is always Person A and Person C getting all €8, as shown in the top figure. In the bottom figure, the Punisher is choosing to reduce €2 of the Wrongdoer and paying €0.20 for that (see the Punisher's earnings in the bottom right). In the experiment, each picture shows 20 different passport-style photographs, which vary based on the condition. For privacy reasons, we are not allowed to display the actual pictures of all study participants in publicly available materials.

EXPERIMENTAL INSTRUCTIONS – WRONGDOER

[English translation; instructions were given in Slovak language]

1. Introduction – when being hired for work

- First, I will ask you to read the consent form, make sure you understand everything and sign it. In case something is unclear, just raise your hand, I will come to you and explain the details.
- You will work for 8 hours. The compensation for your work will be 2.50 euro for each hour, which means20 euro in total. In addition, you may get more money – up to 44 euro extra, depending on the choices of other people.
- The compensation of 20 euros will be paid to you tomorrow after you finish the work. Any additional money, between 0 and 44 euros, will be paid to you in August/September.

2. Wrongdoer decision

[After the work is finished, in two groups of 10-12 subjects]

[After subjects come, check their IDs for name and age (18 or older), give a set of unique ID stickers to each participant.]

Good morning/afternoon/evening, please take a seat. My name is XXX. Today you will go through several tasks that will overall take about 15 minutes. In the tasks, you will be making decisions. Your decisions will stay anonymous; we will never tell anybody how you decided.

Please listen to the instructions carefully. I want you to understand everything before we begin. If something is not clear, raise your hand, I'll come to you, and you can ask questions.

There were people working for us in several other locations in Eastern Slovakia. I will show you a photo of these people from one location other than yours *[show the photo]*.

In this task, you can decide what to do with the 8 euro two different Roma/non-Roma persons from this photo earned for the work. We do not know exactly which persons, but it's for sure someone from the photo. For each person, you can choose one of these five possibilities:

- You can destroy all 8 euro of a person from the photo. The person will get 0 euro.
- You can destroy 6 euro of a person from the photo. The person will get 2 euro.
- You can destroy 4 euro of a person from the photo. The person will get 4 euro.
- You can destroy 2 euro of a person from the photo. The person will get 6 euro.
- You can choose not to destroy any money of a person from the photo and keep his money unchanged at 8 euro.

Your decisions will have real consequences – a person from the photo will get paid as you decide. Your decision is completely anonymous. The person will receive the money based on your decision, but will not receive any further information. He does not see your photo and cannot destroy your income.

Do you have any questions?

OK, now you can decide whether and how much money of a person from the photo you want to destroy. ***[Hand the answer sheets over to the participants.]*** Please take this sheet of paper. First, place your ID sticker in the top right corner of the sheet. In the top part, please circle the option you want to choose for one person from the photo. In the bottom part, circle the option you want to choose for another person from the photo. When you are done, please put the paper in the envelope. I will turn away when you decide, so that you can make your decision in complete privacy.

Thanks for your decision. I will collect the envelopes now.

Now you will have to make two more decisions about 8 euro earned by two persons from this other photo.

As before, for each of the two persons, you can choose to destroy all 8 euro, destroy 6 euro, 4 euro, 2 euro, or you can choose not to destroy anything and keep the person's money unchanged at 8 euro.

Remember that your decisions will have real consequences – a person from the photo will get paid as you decide. Your decision is completely anonymous. The person will receive the money based on your decision, but will not receive any further information. He does not see your photo and cannot destroy your income.

Please take this sheet of paper. *[Hand answer sheets to participants.]* Don't forget to place your ID sticker in the top right corner of the sheet. In the top part, please circle the option you want to choose for one person from the photo. In the bottom part, circle the option you want to choose for another person from the photo. When you are done, please put the paper in the envelope. I will turn away when you decide, so that you can make your decision in complete privacy.

Thanks for your decision. I will collect the envelopes now.

Now I will ask you to fill in a short questionnaire:

3. Questionnaire and conclusion

[Demographic questionnaire]

Age, education, parental education, economic activity (student/employed/unemployed), marital status, number of children, number of people living in the same household, household possessions (car, personal computer/laptop, smartphone, tablet)

I will also ask each of you to proceed to a colleague of mine who will take a photo of yourself.

That's it, thanks for participating! Now I will call you one by one and you will get the 20 euros for yesterday's work. Please wait until you are called.