CERGE
Center for Economics Research and Graduate Education
Charles University Prague

Essays on Discrimination and Endogenous Preferences

Jana Cahlíková

Dissertation

Prague, April 2016

Jana Cahlíková

Essays on Discrimination and Endogenous Preferences

Dissertation

Prague, April 2016

Dissertation Committee

Michal Bauer (CERGE–EI, Charles University; chair)

Randall K. Filer (CERGE–EI, Charles University, Hunter College)

Štěpán Jurajda (CERGE–EI, Charles University)

Peter Katuščák (CERGE–EI, Charles University)

Avner Shaked (CERGE–EI, Charles University, University of Bonn)

Referees

Alexander W. Cappelen (Norwegian School of Economics)

Dmitry Ryvkin (Florida State University)

Table of Contents

Abstract

In the first chapter of this work, I focus on the effects of international mobility on discrimination. Every year, millions of people relocate to a foreign country for school or work. I provide evidence of how international experience shifts preferences and stereotypes related to other nationalities. I use participation in the Erasmus study abroad program to identify the effect of international experience: students who are ready to participate in the Erasmus program are chosen as a control group for students who have returned from studies abroad. Individuals make decisions in a Trust Game and in a Triple Dictator Game. Results show that while students do not differentiate between partners from Northern and Southern Europe in the Trust Game prior to an Erasmus study abroad, students who have returned from Erasmus exhibit less trust towards partners from the South. Behavior towards other nationalities in the Triple Dictator Game is not affected by the Erasmus study experience. Overall, the results suggest that participants learn about cross-country variation in cooperative behavior while abroad and therefore statistical discrimination increases with international experience.

The second chapter concentrates on inter-ethnic interactions. Ethnic hostilities often spread rapidly, making it essential to understand how individual willingness to engage in causing harm is shaped. Here we study the influence of peers among adolescents and present experimental evidence from a region characterized by tensions with Roma, the largest ethnic minority in Europe. We examine the effect of observing choices of randomly assigned peers on the individual willingness to harm majority or Roma counterparts in incentivized tasks. We find that peers are very influential. When choices are performed in isolation or when individuals are exposed to observing the peaceful behavior of peers, subjects do not discriminate against the ethnic minority. In contrast, when subjects are exposed to a peer who harms the ethnic minority instead of to a peer who does not, the likelihood of harming the ethnic minority increases by 60 percentage points and ethnic discrimination emerges. The results are consistent with theories suggesting a parochial response to a threat of ethnic conflict and can help to explain why ethnic hostilities of masses can spread quickly, even in societies with few visible signs of systematic inter-ethnic hatred.

In the third chapter, we study how psycho-social stress affects willingness to compete and performance under tournament incentives across gender. The paper has implications for gender gaps on the labor market, since many key career events involve competition in stressful settings (e.g. entrance exams or job interviews). We use a laboratory eco-

nomic experiment in which a task is compensated under both tournament and piece-rate schemes and subsequently elicit subjects' willingness to compete. Stress is exogenously introduced through a modified version of the Trier Social Stress Test, and stress response is measured by salivary cortisol levels. We find that stress reduces willingness to compete. For female subjects, this can be explained by performance: while tournament incentives increase output in the control group, women under the stress treatment actually perform worse when competition is introduced. For males, output is not affected by the stress treatment and lower competitiveness seems to be preference-based. These results may explain previous findings that men and women react differently to tournament incentives.

Abstrakt

V první kapitole této disertace se zabývám efektem mezinárodní mobility na diskriminaci. Každý rok se miliony lidí stěhují do zahraničí za školou nebo za prací. Tento výzkum ukazuje, jak taková mezinárodní zkušenost mění preference a stereotypy týkající se jiných národností. Pro identifikaci změny využívám účast v programu Erasmus: studenti, kteří se chystají vycestovat do zahraničí jsou bráni jako kontrolní skupina pro studenty, kteří se z programu právě vrátili. Studenti se účastní ekonomického experimentu, kde se rozhodují ve Hře na důvěru (Trust Game) a ve Hře na diktátora (Triple Dictator Game), tak aby bylo možné oddělit změny ve statistické diskriminaci od změn v diskriminaci založené na preferencích. Výsledky ukazují, že zatímco studenti chystající se na program Erasmus nerozlišují ve Hře na důvěru mezi partnery ze severní a jižní Evropy, studenti, kteří se ze studia v zahraničí již vrátili, projevují méně důvěry k partnerům z jižní Evropy. Chování vůči jiným národnostem ve Hře na diktátora se s účastí v programu Erasmus nemění. Celkově výsledky naznačují, že studenti se během programu seznámí s rozdíly v kooperativním chování mezi jednotlivými regiony, a proto zahraniční zkušenost zvyšuje statistickou diskriminaci.

Druhá kapitola se zaměřuje na interakce mezi různými etniky. Etnické konflikty se často velmi rychle šíří, a proto je důležité porozumět, co ovlivňuje ochotu jednotlivce zapojit se do násilného chování. V této kapitole se zabýváme vlivem spolužáků na chování adolescentů v regionu, ve kterém panuje napětí mezi majoritní společností a Romy. Romové tvoří největší etnickou minoritu v Evropě. Pomocí incentivizovaných úloh zkoumáme, jak to, že jednotlivec uvidí rozhodnutí náhodně přiřazeného vrstevníka, ovlivní jeho ochotu poškodit majoritního nebo romského protihráče. Zjišťujeme, že vliv vrstevníků je velký. Pokud se hráči rozhodují samostatně, nebo před svým rozhodnutím vidí, že ostatní se rozhodují mírumilovně, nedochází k diskriminaci vůči etnické minoritě. Pokud ale hráč vidí, že jeho vrstevník poškodil protihráče z etnické minority, jeho vlastní ochota poškodit minoritního hráče vzroste o 60 procentních bodů oproti situaci, kdy vidí vrstevníka chovat se mírumilovně. V této situaci se také objevuje etnická diskriminace. Naše výsledky jsou konzistentí s teorií parochiální reakce na možnost etnického konfliktu a pomáhají vysvětlit, proč se etnické násilnosti mohou rychle rozšířit i ve společnostech, kde nevidíme žádné výrazné známky mezi-etnického nepřátelství.

Ve třetí kapitole zkoumáme, jak psychosociální stres ovlivňuje ochotu mužů a žen soutěžit, a jejich výkonnost v turnajích. V ekonomickém experimentu využíváme úlohu, která je placena jak pomocí turnaje, tak pomocí výplaty za kus, a následně měříme ochotu

účastníků vstoupit do turnaje. Stres je implementován náhodně pomocí modifikované verze protokolu Trier Social Stress Test a stresová reakce je měřena pomocí koncentrace kortizolu ve slinách. Zjišťujeme, že stres snižuje ochotu soutěžit. Pro ženy je tento výsledek vysvětlen výkonností: zatímco u kontrolní skupiny zavedení turnaje zlepšuje výkonnost, u žen ve stresované skupině turnaj výsledky zhorší. U mužů zjišťujeme, že výkonnost není ovlivněná stresem a nižší soutěživost se zdá být daná změnou preferencí pro soutěživé situace. Naše výsledky mohou pomoci vysvětlit předchozí studie ukazující, že muži a ženy různě reagují na soutěživé pobídky v experimentech. Stejně tak jsou důležité pro pochopení genderových rozdílů na trhu práce, vzhledem k tomu, že řada pro kariéru zásadních situací (například přijímací zkoušky, nebo pracovní pohovory) zahrnuje soutěž pod stresem.

Acknowledgments

I am very grateful to everyone who has helped me throughout my PhD studies and with my dissertation work.

First of all, I would like to express my gratitude to my supervisor, Michal Bauer, for showing me the area of economics I find exciting and worth doing research in, and for providing me with all the guidance, support, and encouragement I needed. I am also grateful to other members of my dissertation committee, namely Randy K. Filer, Štěpán Jurajda, Peter Katuščák, and Avner Shaked for their advice and comments that enabled me to improve my research. Thanks also to other professors at CERGE-EI, to members of the ASC, and to the girls from SAO who are always helpful and positive. From external people, I have greatly benefited from discussions with Bertil Tungodden and Alexander W. Cappelen during my visit at NHH in Bergen.

Further, I would like to thank my co-authors, Julie Chytilová, Luboš Cingl, Ian Levely, and Tomáš Želinský, who make our joint work a pleasant experience and from whom I have learned a lot. Big thanks go also to fellow students at CERGE–EI, especially Klára Kalíšková and Vojtěch Bartoš, for teaching me all about positive peer effects. Thanks for many great discussions that helped formulate my research, for providing critical feedback and mainly for always encouraging me.

Finally, I would like to thank my family for always being there for me, for their love, support, and patience.

Financial support from the Czech Science Foundation (grant no. P402/12/G130), the Grant Agency of Charles University (grant no. 1156213), GDN (grant no. RRC 12+46), grant SVV-2012-265 801 and from CEFRES/Komerční banka is gratefully acknowledged.

All errors remaining in this text are the responsibility of the author.

Czech Republic, Prague — Jana Cahlíková
April 2016

Introduction

In my dissertation, I study discrimination and inequality, their stability and possible sources from three different perspectives. The first chapter focuses on discrimination against foreigners and asks whether it can be affected by the degree of exposure to other nationalities. The second chapter discusses the role of social environment on the emergence of discrimination against ethnic minorities. The third chapter studies possible sources of gender inequality on the labor market by examining the role of stress in performance and willingness to compete across gender. I contribute to the literature on discrimination and inequality in the following ways:

The first chapter provides evidence regarding the stability of discrimination against foreigners, evaluating the effects of the Erasmus study abroad program using an economic experiment. The main contribution of this chapter is that it considers changes to both taste-based discrimination (Becker 1971) and statistical discrimination (Arrow 1973; Arrow 1998). The results show that increased exposure to other groups, which is typically considered as a way of reducing discrimination through gaining affective ties or creating a sense of common identity, can strengthen statistical discrimination via a learning channel.

The second chapter, co-authored by Michal Bauer, Julie Chytilová and Tomáš Želinský, contributes to the literature by studying the role of social environment on ethnic discrimination. Using a lab-in-the-field experiment in Eastern Slovakia, we examine the role of peers in the emergence of ethnic discrimination against Roma, who constitute the largest ethnic minority in Europe. Although discriminatory acts often occur when observed by others or jointly with others, economic literature so far has focused on dis-

crimination as an individual choice. As a second distinguishing feature, we consider discrimination in unambiguously hostile behavior, studying the willingness to pay to destroy the money of the counterpart. We show that observing hostile behavior of peers can play a key role in triggering ethnic hostilities.

The third chapter, which is a joint work with Lubomír Cingl and Ian Levely, contributes to the literature on gender differences in willingness to compete. We examine the effect of stress on competitive decisions and performance. Using a laboratory experiment, we can disentangle the effect of stress on performance from the effect of competition on performance and subsequently study self-selection in competitive environments under stress. We show that for women, the combination of stress and competitive incentives (but neither of these factors separately) leads to lower performance and confidence, which in turn results in a lower willingness to enter competitive situations. We argue that as many crucial career events occur under both stress and competitive incentives, our findings could explain some of the gender gap observed on the labor market, and especially the low female representation in highly competitive positions, such as CEOs, politicians, or advanced academic positions.

Chapter 1

Study Abroad Experience and Attitudes Towards Other Nationalities

1.1 Introduction

Even though millions of people relocate abroad every year to study, work, or for personal reasons, little is known about how such experience affects their attitudes towards other nationalities. On one hand, preferences can change through the creation of affective ties or establishment of a sense of common identity; these would predict a decrease in taste-based discrimination (Becker 1971). In fact, increased exposure has long been highlighted as a factor that can help attenuate negative attitudes towards other groups.[1] On the other hand, there is growing literature showing differences in social capital across countries, usually taking interpersonal trust as the primary measure of social capital (Willinger et al. 2003; Holm and Danielson 2005; Buchan, Johnson, and Croson 2006; Guiso, Sapienya, and Zingales 2009; Bornhorst et al. 2010).[2] If people learn about these differences while abroad, statistical discrimination (Arrow 1973; Arrow 1998) towards other nationalities can be magnified by international experience.

[1] This approach is based on the intergroup contact theory (Allport 1954; Pettigrew 1998) and has been applied to tackling discrimination based on gender, race, ethnicity and disability. Empirical evidence shows that changes towards more positive attitudes are indeed possible (Beaman et al. 2009; Clingingsmith, Khwaja, and Kremer 2009; Dobbie and Fryer Jr 2015; Laar et al. 2005; Boisjoly et al. 2006; Pettigrew and Tropp 2006).

[2] (Willinger et al. 2003) compare trust behavior in Germany and France, (Holm and Danielson 2005) Sweden and Tanzania, (Buchan, Johnson, and Croson 2006) compare the U.S., China, Korea and Japan. (Bornhorst et al. 2010) compare European countries, distinguishing between the North and South.

In this paper, I evaluate the effect of a major program aiming to increase the international experience of its participants: the European Union's Erasmus program. I adopt an experimental approach to study attitudes towards other nationalities, which enables me to investigate the learning channel of international experience. I also investigate the effect of international experience on outgroup bias against foreigners, asking whether international experience shifts the sense of common identity.

The Erasmus program is the largest student-exchange program in the world. In total, over 3 million students have participated since its foundation; currently, more than 250,000 students participate annually. To avoid many problems due to selection into the program, I compare students who have just returned from their Erasmus stay to successful applicants who are just about to leave for their stay. I use a Trust Game (Berg, Dickhaut, and McCabe 1995) as a proxy for a "business-like" interaction in which expectations about a partner's behavior play a major role and a Triple Dictator Game as a measure of non-strategic prosocial motivation, following Fershtman and Gneezy (2001) and Bauer, Fiala, and Levely (2014). A total of 199 students from the Czech Republic participated in the experiment, which took place either before or after their Erasmus stay in other European countries. The experiment required them to interact with partners of their own nationality and with partners from other European countries. As an important advantage over studies based on surveys, behavior was incentivized, and participants did not know that they had been invited into the study because of their past or future participation in the Erasmus program.

The main finding of this paper is that, while students do not differentiate between partners from Northern and Southern Europe in the Trust Game prior to Erasmus study, students with Erasmus experience start to exhibit lower trust towards partners from Southern Europe. This discrimination pattern is consistent with the variation in social capital across Europe, and the results overall support the notion that students learn about cross-country differences in cooperative behavior while abroad. In other words, statistical discrimination towards other nationalities seems to become stronger with international experience. As a second finding, Erasmus experience does not shift the bias against foreigners in the Triple Dictator Game, suggesting that the sense of European identity does not increase as a result of the program. However, this is because there is no outgroup bias against foreigners even among students who have not yet studied abroad, plausibly due to (self-) selection.

The existing literature shows that a low level of social capital is linked to the efficiency

of interpersonal interactions within society (Glaeser et al. 2000; Henrich et al. 2001; Alesina and La Ferrara 2002; Henrich et al. 2006; Herrmann, Thöni, and Gächter 2008; Gächter and Herrmann 2011), and therefore can hinder economic development (Knack and Keefer 1997; Tabellini 2010; Gorodnichenko and Roland 2011). My results suggest that when taking a more globalized perspective, low social capital within a society can create additional barriers to development — as people of other nations learn about the low social capital of a country, cross-border interactions can also be affected, including diplomatic negotiations, and the amount of international trade.

Furthermore, this paper contributes to the discussion around group identity and its stability. Group identity plays a major role in interpersonal interactions, due to the potential for discrimination against outgroup members (Akerlof and Kranton 2000). Ingroup favoritism has been identified both among groups created artificially in the laboratory (Tajfel et al. 1971; Charness, Rigotti, and Rustichini 2007; Chen and Li 2009) and among real social groups (Goette, Huffman, and Meier 2006; Bernhard, Fehr, and Fischbacher 2006). Using survey data on trust, Guiso, Sapienya, and Zingales (2009) find that there is ingroup favoritism towards one's own nationality. Exposure to foreigners could, in principle, help create a sense of common identity — a person may become closer to feeling like a "European" or a "world" citizen. Among policy makers, there is much optimism regarding this channel.[3] Unfortunately, there is little evidence to support these claims. The main problem is that most studies do not separate the effects of the program from the selection effect.[4] Selection into the Erasmus program is an important issue, as under current conditions only about 5% of all European students participate in the program.

My results suggest that the strength of European identity does not change as a result of Erasmus experience. If there is a shift in the sense of European identity, ingroup favoritism towards one's own nationality should diminish or disappear. But I do not find any bias against other nationalities in the Triple Dictator Game for the students before or after their Erasmus stay. The two samples also respond similarly when asked about the strength of European identity in a questionnaire. In this highly selective environment,

[3]The Erasmus program proclaims to be "changing lives, opening minds", and believes that "[T]heir experiences give students a better sense of what it means to be a European citizen." Source: http://europa.eu/youth/article/erasmus-exchange-programme_en

[4]See the literature survey in Di Pietro (2015). The few exceptions focus on labor market outcomes: Parey and Waldinger (2010) and Di Pietro (2015) find a positive effect of a study abroad stay on future international labor mobility and employability, respectively, using a variation in program availability as an instrument for the participation decision. The European Commission has only recently published an evaluation of Erasmus that acknowledges the problem of selection and partially implements an ex-ante/ex-post survey design. See European Commission (2014).

students selected for the program seem to feel quite European even before their Erasmus stay. Evaluating the issue of selection further, students in my sample who do not intend to go on Erasmus are more biased against foreign partners and feel less European. Therefore, it seems the popular view that the Erasmus program strengthens the sense of European identity is driven by the selection into the program and not by the effects of the program. Still, the effects of Erasmus estimated in this article should be viewed as the average treatment effect on the treated. Potentially, if the program were able to target students who feel less European to begin with, there would be room for the "common identity building" channel of international experience to operate.

1.2 Experimental design

To identify the effect of international experience on preferences and stereotypes towards other nationalities, the research design consists of an experiment run on specific subject pools that differ in their degree of international experience — students before and after an Erasmus study-abroad stay. This section first describes the sample selection and then presents details of the experiment procedure.

1.2.1 Sample selection

This paper uses Erasmus program participation as the source of variation in international experience. I use a between-subject design. Successful applicants who were just about to go on their Erasmus stay at the time of the experiment were taken as a control group for students who had just returned from their Erasmus stay. The experiment took place in Prague, the Czech Republic, and the sample selection process can be summarized as follows:

I cooperated with the largest university in the Czech Republic, Charles University in Prague,[5] and obtained a database of all their students who were enrolled in the Erasmus program in the academic year 2011/12 (1009 students) and in the academic year 2012/13 (923 students). Students from the 2012/13 database were recruited as "Before Erasmus" subjects for sessions that took place in June 2012, while students from the

[5]Charles University has over 50,000 registered students. It also sends more students to the Erasmus program than any other Czech school; for illustration, 5,589 students from Czech universities participated in the Erasmus program in the academic year 2010/2011. Out of these, almost one fifth (1,056), were from Charles University.

2011/12 database were recruited as "After Erasmus" subjects for sessions in June 2012 and November 2012.[6]

The email invitation to the experiment did not mention the Erasmus program, but encouraged the recipient to take part in a paid experiment in decision making. The e-mail included a personalized link, which was used for online registration into one of the available sessions. Overall, more slots were opened for the "After Erasmus" students compared to the "Before Erasmus" students and more students from the 2011/12 database were invited, compared to the 2012/13 database. This is because the "After Erasmus" had to be invited to both June 2012 and November 2012 sessions, to allow a control for time effect; see the discussion in section 2.2.

The two main samples consist of 75 local students who were about to leave on their Erasmus stay in other European countries ("Before Erasmus" sample) and 124 local students who had already returned from their study-abroad stay ("After Erasmus" sample).[7] Summary statistics of the "Before Erasmus" and "After Erasmus" samples are presented in Table 1.1, which shows that the two samples do not differ in characteristics other than age. There is a sufficient variation in terms of age when students go on Erasmus, so age can and will be controlled for in the analysis.[8]

Apart from the two main samples "Before Erasmus" and "After Erasmus", there are two auxiliary samples:

First, the aim is to study behavior towards partners of different nationalities, and to do so without deception. Therefore, international students had to be recruited. Incoming Erasmus students at Charles University were invited by e-mail, and a further recruitment campaign was run on social networks. To ensure sufficient variety of nationalities dur-

[6]Each of the 2011/12 and 2012/13 populations was divided into thirds using stratified random sampling, with stratification based on gender, study major and the region of the Erasmus stay. Two thirds of each population were invited for participation in the experiments in June 2012; two thirds of the 2011/12 database were invited again in November 2012. This means that one-third of the 2011/12 population was invited twice; however, each subject could participate only once.

[7]Both Czech and Slovak students are perceived as local in the baseline analysis. Slovak students are largely present at Czech universities, due to the lack of a language barrier and cultural proximity. For Charles University, 13.7% of students are foreigners, of which Slovak students form 46%, according to the 2011 annual report. The results presented in the text are robust to being limited to Czech subjects only. Also, one subject about to go on Erasmus and three subjects with Erasmus experience are neither Czech nor Slovak, but are foreign students doing their degree in Prague and going on Erasmus elsewhere. These subjects are not included in the baseline "Before Erasmus" and "After Erasmus" samples. However, the results presented below are robust to including these subjects.

[8]Furthermore, several robustness checks were performed to make sure the effect of study-abroad stay is estimated, not the effect of age, such as restricting the sample to common support in terms of age. Results are available upon request.

ing the experiment, each session had hidden registration limits for local subjects and international subjects, where the limits were set separately for subjects from Northern and Southern Europe. Overall, 126 international students from Northern and Southern Europe participated in the experiment.[9]

Second, a sample of 53 local students with no connection to the Erasmus program ("Never Erasmus" sample) is used to consider selection into the program. These students were recruited through the social network campaign and their Erasmus status was checked using the database of all Erasmus stays in the past years and by asking questions about study-abroad experience in the end-questionnaire.

1.2.2 Identifying assumptions

For the identification strategy to hold, three assumptions must be made:

First, the pools of students going on Erasmus in the two consecutive years 2011/12 and 2012/13 must be the same, in terms of baseline attitudes towards other nationalities. In other words, the only difference between the two pools is the realized stay abroad. The Erasmus program did not change between the two academic years, nor did the selection processes. Comparing the observable characteristics of the 2011/12 and 2012/3 databases of all outbound Charles University students, there are no significant differences between the two pools in terms of gender, field of study, or the region of the Erasmus stay (see columns 1 and 2 of Appendix Table 1.9). The only difference is that more students in the 2012/13 database were enrolled in a BA-level program at the time of application.

The second assumption is that preferences towards specific nationalities did not change between June 2012, when "Before Erasmus" students participated in the experiment, and November 2012, when most "After Erasmus" students participated.[10] This is the reason why some "After Erasmus" students were invited into the June 2012 sessions — a robustness check can be run by comparing the two "After Erasmus" subsamples.

[9] A smaller number of slots was opened for students of other nationalities, to avoid suspicion regarding the purpose of the research project during recruitment and during the experiment itself. A total of 38 international students from countries outside Northern and Southern Europe participated in the experiment. These students came from Bulgaria, Georgia, Hungary, Lithuania, Macedonia, Russia, Slovenia, Turkey, Ukraine, USA and Vietnam. As there is no clear prediction regarding changes in behavior of local (Czech and Slovak) students towards subjects from these countries following a study abroad stay in Northern and Southern Europe, these observations are excluded from the analysis. Still, the results presented in the paper are robust to including these observations.

[10] It was impossible to run all sessions in the same month — many 2011/12 outbound students were not back from their stay by June 2012, while many 2012/13 outbound students would be gone by September 2012.

Third, and most importantly, the experiment participants "After Erasmus" and "Before Erasmus" cannot differ in aspects other than the international experience itself, i.e. recruitment from the 2011/12 and 2012/13 databases of outbound Erasmus students must be equally successful. The recruitment process consisting of e-mail invitations and online registration was described above and was identical for the two pools. Most slots opened for registration were filled and the response rates were similar for the 2011/12 and 2012/13 databases — in respect to the number of experiment participants in relation to the number of invitations sent, the response rates are 11.8% and 12.1%, respectively. I have already argued that the two samples do not differ in characteristics other than age (see Table 1.1).

Last but not least, the experiment samples "Before Erasmus" and "After Erasmus" can be compared to all Charles University outbound Erasmus students in the respective years. See Appendix Table 1.9. Considering the characteristics available (gender, level of study, field of study, host country), recruitment into the experiment seems to be successful. There are 10% more males than would be typical in the program, more students of Business, Economics and Law, and fewer students of Medicine. However, these differences can potentially be attributed to the gender limits set in recruitment.[11] Gender variety was needed for the chosen design which manipulates nationality, gender, and field of study of game partners. The gender limits were more likely to be binding for females, as women form a vast majority (around 70%) of all Erasmus program participants.

1.2.3 Experimental procedure

Seventeen experiment sessions were organized (nine in June 2012 and eight in November 2013), with the number of subjects per session ranging from 20 to 28. All sessions took place at the Laboratory of Experimental Economics in Prague. Each session consisted of an introduction in which participants recorded their nationality, gender, age and study major, followed by the main section in which the Triple Dictator Game and the Trust Game were played in a randomized order, of a payoff stage where the individual payoffs were determined, and of an end-questionnaire that focused primarily on the past international experience of the subjects. The experiment was programmed and conducted using the software z-TREE (Fischbacher 2007).

Participants received written instructions before each stage of the experiment. All

[11] The gender ratio in the experimental sessions could not exceed two thirds in either direction.

payoffs were stated in experimental currency units (ECU). Participants did not receive any feedback on their performance or payoff until the final stage, where they randomly (by hitting buttons on the screen) selected decisions relevant for payment. The experiment lasted on average 2 hours and the average payment was CZK 457 (approximately EUR 18).[12]

It is important to note that subjects' Erasmus program (past or future) participation was not mentioned in the invitation or at any point during the experiment.[13]

Experimental tasks

In the Trust Game, Player A ("Sender") had an endowment of 100 points, while Player B ("Receiver") had an endowment of 0. In the first stage, Player A decided whether and how much s/he wished to transfer to Player B, choosing between 0, 20, 40, 60, 80 and 100 points. The amount sent was tripled. In the second stage, Player B decided how many points s/he wanted to send back to Player A for any amount potentially sent by Player A, i.e. a strategy method was used. The structure of the game was common knowledge. In addition to actions, beliefs were also elicited. Specifically, these were Player A's first-order and second-order beliefs and Player B's first order beliefs.[14]

The structure of the Triple Dictator Game is similar to the Trust Game, except that there is no second stage. Player A decided whether and how much s/he wished to transfer to Player B, choosing between 0, 20, 40, 60, 80 and 100 points and the amount sent was tripled. However, Player B was only a passive receiver of Player A's points and did not make any active decisions. S/he was asked to report only his/her first-order beliefs, i.e. how much s/he thought Player A would send. Player A's second-order beliefs were also elicited.

Each subject played both roles, Player A and B. The order of roles was randomized across sessions, and subjects learned of the existence of the second part only after they finished their decisions in the first role.

[12]Student wages in Prague are around EUR 3-4/hour on average.

[13]At the end of the experiment, students were asked to state the perceived purpose of the study. Erasmus program participation was not mentioned by any subject.

[14]How much Player A thinks B will return for the amount actually sent, how much Player A thinks B expects from him, and how much Player B expects from A, respectively. Subjects receive a bonus of 20 points if they guess correctly. One round is chosen randomly for the payment on beliefs and one partner from that round is relevant for payment.

Manipulating a partner's characteristics

The identity of partners was varied on a within-subject level. In each game, Player A was asked to make decisions about sixteen potential Player Bs. Each partner was characterized by a profile stating nationality, gender, age, and field of study.[15] Analogously, Player B was asked to make decisions regarding sixteen potential Player As. The decision maker always saw four profiles of potential partners at once and played four of these rounds. To determine the composition of partners' profiles in a given round, session participants were randomly matched in groups of four and one hypothetical profile was added.[16] The profiles were displayed in a random order. One of the sixteen decisions in each role was relevant for payment.

In this paper, a partner's nationality is of primary interest. Additional information was used to decrease the risk of an experimenter-demand effect (Bardsley 2005), while ensuring that nationality was sufficiently salient. Limits set during the registration process ensured enough variation in nationalities and gender within each session.

The Trust Game was applied in the above setting to study how trust behavior is influenced by the partner's nationality. Behavior in the Triple Dictator Game can be used as a measure of non-strategic prosocial preferences, jointly capturing preferences for altruism, inequality aversion, and efficiency maximization. Therefore, observing behavior in the Triple Dictator Game can help to disentangle preference-based and beliefs-based components of trust.

1.3 Results

1.3.1 Learning channel of international experience

I first explore whether students learn about cross-country differences in values and behavior while abroad. To test this "learning channel", I examine how senders before and after an Erasmus stay differentiate between partners from Northern and Southern Eu-

[15] Participants were asked to provide nationality, gender, age, and field of study at the beginning of the experiment and knew this information would be displayed to the decision makers. Five categories were distinguished with respect to field of study: Business, Economics or Law; Humanities, Social Sciences or Education; Math, Physics, Natural Sciences or Technical; Medicine; Arts, Philosophy and Languages.

[16] The hypothetical profile, which was the same for all subjects in a given round, was added to ensure enough variation in partner profiles. No deception was involved as players were always asked to state their decisions for all four potential partners they could see, but knew that they would be matched with only one of the four.

rope. The choice of these two regions is motivated by the variation in social capital across Europe. Focusing on interpersonal trust as the principle measure of social capital, people from Southern Europe are much less likely to state that other people can be trusted, compared to people from Northern Europe; see Figure 1.1.[17] I hypothesize that with a study abroad experience, students learn about differences in social capital across Europe and start to differentiate more between partners from Northern and Southern Europe. This effect should be more pronounced in the Trust Game, where a partner's behavior actually matters. The division of countries into Northern and Southern Europe as used in the analysis is presented in Table 1.2.

Trust Game - partners from Northern vs. Southern Europe

Mean behavior in the Trust Game by the Erasmus status of the sender and by the nationality of the receiver is presented in Panel A of Table 1.3. I will focus on discussing the average amounts sent towards partners from Northern and Southern Europe, where I have a clear prediction regarding the direction of the change due to learning. Figure 1.2 presents the results in levels (Panel A) and as a difference in behavior between partners from the two regions (Panel B). Senders "Before Erasmus" sent on average 56.9 points to partners from the North and 60.1 points to partners from the South. This means that they felt actually more favorable towards Southern receivers, but the difference is not significant (Wilcoxon rank-sum test, $p = 0.321$). Senders "After Erasmus", on the other hand, sent significantly more points to Northern partners than to Southern partners (58.1 vs. 52.8 points, $p = 0.029$). Put differently, while subjects from Northern Europe received similar amounts from senders before and after Erasmus ($p = 0.652$), subjects from Southern Europe received significantly lower amounts from senders with more international experience ($p = 0.019$).

So far, the results have shown that students "Before Erasmus" do not discriminate between partners from Northern and Southern Europe, while students "After Erasmus" do. Next, I test whether the discrimination pattern changes with a study abroad experience, using a regression analysis.

The following regression model is estimated:

[17] Data from the World Values Survey (WVS) are used. The Figure summarizes answers to the WVS question "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?" See Appendix Figure 1.5 for a more detailed map of trust across European countries.

$$AmountSent_{i,j}^{TG} = \alpha + \beta AfterErasmus_{i,j} + \gamma RecipientSouth_{i,j} + \delta RecipientSouth_{i,j} * AfterErasmus_{i,j} + X'_{i,j}\theta + \varepsilon_{i,j} \quad (1.1)$$

, where $AmountSent_{i,j}^{TG}$ is the amount of points sent in the Trust Game by sender i to receiver j. The Erasmus status of the sender is captured by an indicator variable $AfterErasmus_{i,j}$ and the nationality of the receiver by an indicator variable $RecipientSouth_{i,j}$. The baseline is therefore the amount sent by senders "Before Erasmus" to partners from Northern Europe. Vector $X_{i,j}$ consists of a range of other control variables described below. The interaction term $RecipientSouth_{i,j} * AfterErasmus_{i,j}$ is of primary interest. The coefficient δ captures how the discrimination pattern between Northern and Southern partners changes as a result of Erasmus program participation. Standard errors are clustered on the sender level.

Estimation results are presented in Table 1.4 and confirm that discrimination between Northern and Southern partners changes significantly with a study abroad experience — the negative effect of an Erasmus stay on the amount sent is specific for Southern partners (column 1, $p = 0.044$). This result holds when controlling for other senders' and receivers' characteristics observable through the games (gender, age, field of study), for the order of the two games (Trust Game, Triple Dictator Game), and for the order of the two roles (sender, receiver); see columns 2-3 of Table 1.4.

Result 1: An Erasmus study abroad stay changes how students discriminate between partners from Northern and Southern Europe in the Trust Game. While students prior to Erasmus study abroad do not differentiate between partners from the two regions, students with study abroad experience send lower amounts to partners from Southern Europe.

I next perform three robustness checks with respect to Result 1:

First, Equation 1 is estimated using ordered probit instead of OLS, to take into account the discrete nature of the dependent variable. Estimation results are presented in Appendix Table 1.10. Students after an Erasmus study abroad are significantly less likely to send 100 and 80 points to Southern partners, and more likely to send 0, 20, and 40 points to Southern partners, confirming that there is a negative effect of Erasmus program participation on behavior towards Southern partners.

Second, I add behavior towards local (Czech and Slovak) partners into the picture. As is visible from Table 1.3, behavior towards local partners in the Trust Game is not affected by an Erasmus stay abroad (Wilcoxon rank-sum test, $p = 0.677$). Therefore, an Erasmus stay has a negative impact only on behavior towards partners from Southern Europe, while it does not affect behavior towards local partners or partners from Northern Europe. In other words, while subjects before an Erasmus study abroad did not differentiate at all based on a partner's nationality in the Trust Game, subjects after an Erasmus study abroad behave less favorably towards partners from Southern Europe, compared to local partners or compared to partners from Northern Europe.

Third, I run a check showing that the estimated effect is not driven by changed preferences/beliefs regarding Southern partners between June 2012 (when "Before Erasmus" students participated in the experiment) and November 2012 (when most "After Erasmus" students participated). Appendix Figure 1.6 presents how senders differentiate between partners from Northern and Southern Europe, splitting the "After Erasmus" sample into June 2012 and November 2012 participants. If anything, the negative effect towards Southern partners is stronger among "After Erasmus" subjects who participated in June 2012.

Effect by the region of Erasmus study abroad

It is important to note that students going on Erasmus necessarily meet people from both Northern and Southern Europe, regardless of where they go. However, the learning effect of international experience can still differ by the region of the study abroad stay. I investigate this possibility by re-estimating the effect of Erasmus experience on behavior in the Trust Game separately for students with a (planned or realized) Erasmus stay in Northern Europe and separately for students with an Erasmus stay in the South.

Regression results are presented in Table 1.5. The effect goes in the same direction for students going abroad to Northern and Southern Europe, but the strength of the effect and the underlying story differ. Subjects going "North" (column 2-4 of Table 1.5) do not differentiate between partners from Northern and Southern Europe before their Erasmus stay, but they send significantly less to partners from the South after their stay ($p = 0.031$). The effect of Erasmus on discrimination between Northern and Southern partners, as captured by the variable $AfterErasmus * ReceiverSouth$, is negative, not significant when focusing on this subgroup separately ($p = 0.271$).

Subjects going "South" (column 5-7 of Table 1.5) show a strong preferential treatment of partners from Southern Europe before the Erasmus stay, suggesting self-selection in terms of where students decide to go — holding a positive image of Southern Europe, the students decide to go "South". The Erasmus study abroad experience then dramatically changes how students differentiate between partners from Northern and Southern Europe. Students with Erasmus experience show higher trust towards partners from Northern Europe, even though this difference is not significant. The change in discrimination pattern with Erasmus, as captured by the variable $AfterErasmus * ReceiverSouth$, is strong and significant ($p = 0.070$). In terms of effect size, students with experience in Southern Europe are driving the overall negative effect of Erasmus on trust towards Southern partners.

Beliefs about partners from Northern vs. Southern Europe

To attribute the observed changes in relative behavior towards partners from Northern and Southern Europe to learning about differences in social capital across countries, I next examine the two measures of beliefs elicited during the experiment.

First, I focus on beliefs about expected trustworthiness, defined as the expected amount returned by the receiver, in % of what was sent to the receiver. Note that the measure of expected trustworthiness is potentially problematic as senders were asked how much they think Player B would return only for the amount that was actually sent. As subjects "After Erasmus" actually sent lower amounts to partners from the "South", the expected trustworthiness is elicited for amounts sent that were on average lower (plus beliefs about trustworthiness are not elicited for subjects who sent 0 points to the receiver). Still, as Panel A of Figure 1.3 shows, beliefs about trustworthiness of Northern versus Southern partners move in the direction that corresponds to the change observed in the Trust Game, but the beliefs are quite noisy and the change is not statistically significant ($p = 0.374$).

As a cleaner measure of the change in beliefs, I next examine beliefs about trust behavior of senders from Northern and Southern Europe.[18] [19]

[18] Beliefs about points received from these senders in the Trust Game from the position of local receivers before or after their Erasmus stay.

[19] Trust and trustworthiness behavior are closely linked. When considering individual-level behavior in my sample, trust and trustworthiness behavior is significantly correlated, both for the local students (Spearman's rank correlation, $\rho = 0.4832, p < 0.001$) and for foreigners from Northern Europe ($\rho = 0.4968, p < 0.001$) and Southern Europe ($\rho = 0.4621, p < 0.001$). Appendix Figure 1.7 summarizes this result graphically. As a measure of an individual's trust level, I computed the average

The effect of Erasmus on beliefs about partners' trust behavior is presented in Panel B of Figure 1.3. The change in beliefs about trust mirrors the effect found for the beliefs about trustworthiness (Panel A), but the effect is stronger and statistically significant ($p = 0.005$). While subjects "Before Erasmus" expect senders from Southern Europe to send more in the Trust Game compared to Northern senders, subjects "After Erasmus" expect senders from Southern Europe to be less trusting than Northern senders. The latter pattern is consistent with the map of interpersonal trust across Europe shown in Figure 1.1. Results from regression analysis are presented in Appendix Table 1.11 and confirm that the change in beliefs regarding trust behavior of Southern partners is large and statistically significant even when controlling for the observable characteristics of senders and receivers, and for order effects.

Triple DG - partners from Northern vs. Southern Europe

Amounts sent in the Triple Dictator Game are presented in Panel B of Table 1.3 and in Figure 1.4. While "Before Erasmus" students treat Southern partners more favorably than Northern partners in the Triple Dictator Game (sending 31.3 vs. 28 points to the two groups, $p = 0.131$, Wilcoxon rank-sum test), this difference disappears after Erasmus study abroad. Students "After Erasmus" actually send more points to Northern partners, but the difference is small and insignificant (24.4 points vs. 26 points, $p = 0.506$).

Estimation results then show that an Erasmus stay has a negative impact on the amount sent to Southern partners relative to Northern partners (column 1 of Table 1.6), but the effect is significant at 10% level only when controlling for additional characteristics (columns 2-3). The effect thus goes in the same direction as the effect in the Trust Game, but is weaker. More importantly, when adding local partners to the picture, it is clear that the negative effect of Erasmus on the amount sent in the Triple Dictator Game is not unique to receivers from Southern Europe (see Panel B of Table 1.3). Rather, there is a general negative effect of Erasmus program participation on the amount sent in the Triple Dictator Game and the response does not significantly differ between local and Southern or Northern partners, as shown in Table 1.7 (columns 4-6). This is in comparison to the behavior in the Trust Game, where the negative effect of the Erasmus

amount sent in the Trust Game, averaging over the 16 profiles of potential partners. As a measure of an individual's trustworthiness, I computed the average return ratio (*Return ratio= amount returned to sender/(3*amount sent by sender)*, averaging over all receivers' decisions. Each receiver makes 80 trustworthiness decisions — for 16 profiles of potential senders and 5 trustworthiness decisions per sender, as receivers' decisions were elicited using the strategy method.

program participation was observed only for partners from Southern Europe.

Behavior in the Triple Dictator Game can be used to measure non-strategic prosocial motivations towards partners from Northern and Southern Europe. As these motives can be present also in the Trust Game, I want to test whether the observed changes in the Trust Game are caused by the preference-based component of trust or the beliefs-based component of trust. This is done by re-estimating the effect of Erasmus on discrimination between Northern and Southern partners in the Trust Game, this time controlling for behavior towards these partners in the Triple Dictator Game.

Estimation results are present in columns 4-6 of Table 1.4. The negative effect of Erasmus study abroad on behavior specifically towards partners from Southern Europe in the Trust Game persists even when controlling for the behavior in the Triple Dictator Game. These results suggest that the differentiation between partners from Northern and Southern Europe in the Trust Game among subjects with study abroad experience (as presented in Panel A of Table 1.3) cannot be explained by differences in the preference-based, non-strategic component of trust.[20]

Discussion

Overall, the results from this section show that students with international experience start to differentiate between partners from Northern and Southern Europe. Presumably, this effect is driven by learning about behavioral differences across regions while abroad. Linking the results to different sources of discrimination, it seems that it is the statistical discrimination which emerges with increased international experience. There are four main arguments for such a claim:

First, students with study abroad experience start to differentiate between partners from Northern and Southern Europe in the Trust Game, in which the expected behavior of partners actually matters. On the contrary, students after a study abroad program do not discriminate based on a partner's nationality in the Triple Dictator Game, where expectations about a partner's behavior do not matter. Changes in observed behavior

[20]Note that the behavior in the Triple Dictator Game is significantly correlated with the behavior in the Trust Game ($p < 0.01$), but the estimated coefficient is significantly below 1. Specifically, the point estimate lies between 0.4 and 0.5 for all samples of local (Czech and Slovak) students — those "Before Erasmus","After Erasmus" and also for Erasmus non-participants ("Never Erasmus"); detailed results available upon request. Moreover, I cannot reject the null hypothesis that the coefficient is the same across the three groups of local students. This suggests that while the non-strategic motives indeed matter for the decision in the Trust Game, the "business-like" setting of the game crowds-out the prosocial motivations present in the Triple Dictator Game.

in the Trust-Game seem to be driven by the beliefs-based component of trust, as they persist even when controlling for the preference-based motives using the Triple Dictator Game.

Second, the way students with Erasmus experience differentiate between partners in the Trust Game is consistent with the variation in social capital across Europe. Southern Europe scores much lower in interpersonal trust than Northern Europe (see Figure 1). While Czech students with less international experience ("Before Erasmus") do not differentiate between Northern and Southern partners in the Trust Game, students after Erasmus study abroad show lower trust towards Southern partners, possibly because they learned about low social capital in the South.

Third, while the effect on behavior towards Southern partners goes in the same direction for students who went on a study abroad to Northern Europe, the effect is stronger among subjects with study abroad experience from Southern Europe, who are more likely to encounter behavioral differences in the South.

Fourth, the observed change in behavior in the Trust Game is accompanied by a change in beliefs regarding Northern and Southern partners, which were measured separately.

Behavior of foreigners

The behavior and beliefs of subjects who returned from an Erasmus study abroad is consistent with the explanation that they learned about relatively lower social capital in Southern Europe while abroad. However, this may not correctly reflect the behavior of Northern and Southern subjects in my sample. The foreigners from Northern and Southern Europe who took part in the experiment — mostly Erasmus students studying in Prague — are not by any means a representative sample of students from these regions. The program is very selective in general. Moreover, these students chose to study in Prague.

In this subsection I examine whether Southern students in my sample (N=78) are less trustworthy than Northern students in my sample (N=45). I computed individual-level trustworthiness as the average return ratio (*Return ratio= amount returned to sender/(3*amount sent by sender*), averaging over all senders and all strategy method levels. I find that there is no significant difference in trustworthiness between individuals from the two regions. Northern receivers send back on average 21% of the amount received, while Southern receivers return on average 23% (Wilcoxon rank-sum

test, $p = 0.395$).[21]

Focusing instead on trust behavior (calculated for each individual as the average amount sent across all partners), I find that Northern students are more trusting than Southern students, sending on average 55.1 vs. 48.4 points, but the difference is not statistically significant ($p = 0.240$). Interestingly, while Southern senders do not differentiate between Czech, Northern and Southern partners, Northern senders send significantly lower amounts to partners from Southern Europe, relative to Czech or Northern partners, see Appendix Table 1.12.

There are two main messages from this subsection. First, the behavior and beliefs of subjects "After Erasmus" seems to respond to relatively lower social capital in Southern Europe compared to Northern Europe. However, subjects seem to take this approach even towards a very selected group of foreigners studying in Prague, among whom we do not observe significant regional differences in trust or trustworthiness behavior. Discrimination between Northern and Southern partners in the experiment can therefore be interpreted as statistical discrimination based on incorrect beliefs. Whether these beliefs would be correct for a more representative sample of Northern and Southern students we do not know.

Second, the fact that Northern students in my sample also discriminate against Southern partners in the Trust Game suggests that the learning channel of international experience can go in two ways: i) students learn about the behavioral differences in Southern Europe while abroad and ii) students learn how people from Northern Europe perceive Southern Europe while abroad.

1.3.2 International experience and bias against foreigners

After examining the learning channel of international experience, this section focuses on the change in preferences towards foreigners. The logic of the analysis performed here is different from that used in the previous section. The hypothesis is that the study-abroad experience leads to more favorable treatment of foreign partners through creating affective ties or through strengthening a sense of common identity. Therefore, I no longer compare behavior towards Northern and Southern partners. Rather, I examine

[21] Separating the trustworthiness for different levels of senders' trust (sending 20, 40, 60, 80, or 100 points), Southern partners are relatively more trustworthy for low levels of senders' trust. They return on average 13% of the amount received if senders send 20 points, compared to 9% among Northern receivers ($p = 0.132$), and 20% vs. 16% if senders send 40 points, ($p = 0.228$).

how students behaved towards their ingroup (partners of students' own nationality) and outgroup (partners of other nationalities) and whether the ingroup favoritism diminished with an Erasmus stay. The focus is on the behavior in the Triple Dictator Game — it provides a cleaner measure of non-strategic prosocial preferences, as the partner has only a passive role in this game.

Triple Dictator game - local vs. foreign partners

Panel B of Table 1.3 presents the average amounts sent in the Dictator Game by the Erasmus status of the sender ("Before Erasmus" or "After Erasmus") and by whether the partner was local or foreign. The results show that there is only small and insignificant ingroup favoritism towards their own nationality for the students who were about to participate in the Erasmus program (Wilcoxon rank-sum test, $p = 0.829$). Study-abroad experience then has a negative impact on the amount sent in the Triple Dictator Game, both when the partners are of the sender's own nationality ($p = 0.075$) and when they are of a different nationality ($p = 0.004$). The in-group favoritism among students "After Erasmus" is still rather small and insignificant ($p = 0.154$).

Estimation results presented in columns 1-3 of Table 1.7 confirm that the discrimination pattern between local and foreign partners does not change with Erasmus (as captured by the variable $ReceiverForeign * AfterErasmus$). When controlling for other characteristics, subjects after an Erasmus stay sent lower amounts than subjects before Erasmus, but neither group significantly differentiates between local and foreign recipients. After disentangling international partners from Northern and Southern Europe (columns 4-6 of Table 1.7), the results show that among students "Before Erasmus", there is a small significant bias against partners from Northern Europe, when controlling for other characteristics. Potentially, students "Before Erasmus" perceived partners from the North as wealthier and therefore less needy than other partners.

Overall, the results of this experiment do not support the hypothesis that international experience lessens negative attitudes towards foreigners. This is primarily because no preferential treatment of the subjects' own nationality was found for students prior to an Erasmus study abroad. If there is any effect at all, then the ingroup favoritism is slightly greater for students with study-abroad experience.

Result 2: An Erasmus study abroad stay does not change how students discriminate between local and foreign partners in the Triple Dictator Game. No outgroup bias against

foreign partners was found and this result holds both for the students who are about to leave for their Erasmus stay, and for those who have already returned.

Self-selection into studies abroad

There are two possible explanations for Result 2. Either there is no ingroup favoritism towards their own nationality in the population of students, or those who self-select into going abroad already identify as "European", which is why they do not distinguish between partners of their own nationality and foreign partners. My results provide suggestive evidence for the latter argument.

To investigate the role of (self-) selection, I compare the behavior of students "Before Erasmus" to an auxiliary sample of 53 non-participants (the "Never Erasmus" sample).[22] Estimation results are reported in Table 1.8. The "Never Erasmus" students differentiate more between local and foreign partners in the Triple Dictator Game than students "Before Erasmus". The outgroup bias is about twice the size and the null hypothesis of no outgroup bias can be rejected at the 5% level for the "Never Erasmus" students ($p = 0.043$), when controlling for other observable characteristics (columns 1-3).

Using data from the end-questionnaire, the "Never Erasmus" students are less likely to identify themselves as part of the European Union, compared to students "After Erasmus" (58.6% vs. 83.1% $p = 0.001$, Wilcoxon rank-sum test). If not taking selection into account, one could mistakenly conclude that international experience inspires students to identify more strongly as European. However, including the "Before Erasmus" students into the picture shows that there is a large and statistically significant difference between the non-participants and students who are about to participate in the program (58.6% vs. 80%, $p = 0.009$). The effect of the program — a difference between the "Before Erasmus" and "After Erasmus" students — is only small and statistically insignificant (80% vs. 83.1% who claim to feel they are members of the European Union, $p = 0.197$). This further highlights the advantages of the identification strategy used in this paper.

If one of the intentions of study-abroad programs is to create a sense of common identity, the results of this research suggest that the programs should try to recruit more students and especially target those who feel less "international" to begin with. Also, it may be worthwhile targeting younger students, as results from behavioral studies show that the most sensitive window for the formation of individual preferences and group-

[22] The results should be perceived as suggestive evidence only, because I cannot claim that these students are a representative sample of all students who do not participate in the Erasmus program.

identity occurs at an earlier age — during childhood and adolescence (Almas et al. 2010; Fehr, Glätzle-Rützler, and Sutter 2013; Bauer et al. 2014).

1.4 Conclusion

This paper examines whether attitudes toward other nationalities change with international experience. The variation in international experience was obtained by exploiting student participation in the Erasmus study-abroad program — the behavior of students who were about to participate in the program (75 students) was compared to that of students who had already completed their study abroad stay (124 students). Participants anonymously interacted with partners of different nationalities in a Trust Game and in a Triple Dictator Game. The Triple Dictator Game was used to control for a preference-based component of trust, helping to disentangle statistical and taste-based discrimination.

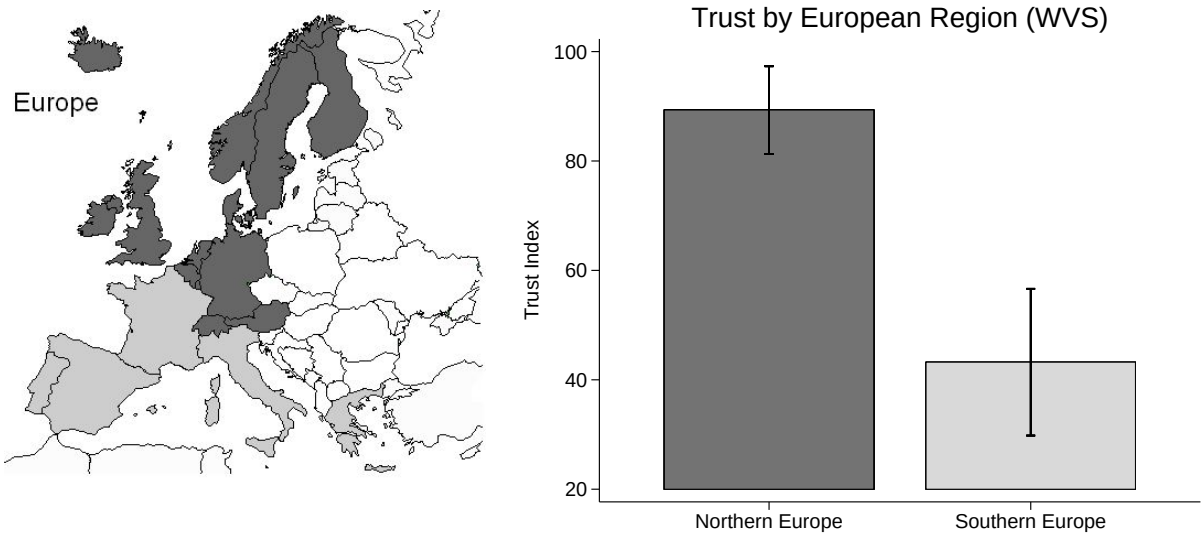
I found the study-abroad experience affected behavior towards other nationalities, and specifically so in the Trust Game. While subjects prior to an Erasmus stay did not differentiate between partners from Northern and Southern Europe, subjects with study-abroad experience started to do so, exhibiting lower trust towards partners from the South. This result holds even when controlling for behavior in the Triple Dictator Game. Such a discrimination pattern is consistent with the lower rank of Southern countries in terms of general trust, as measured by the World Values Survey. As there is also an accompanying change in beliefs about cooperative behavior of partners from Southern Europe, the results overall support the hypothesis that people learn more about cross-country differences in social capital while abroad and subsequently change their behavior according to their experiences. Therefore, the results suggest that statistical discrimination towards different nationalities increases with international experience. This means that in a situation where there are differences in social capital across countries, globalization can create additional challenges for countries with lower social capital.

Examining next whether international experience changes preferences towards foreigners overall, I focused on behavior in the Triple Dictator Game and examined the strength of ingroup favoritism towards partners of a student's own nationality. The results show that even before their Erasmus stay abroad, senders do not show preferential treatment of partners coming from the same country, and preferences towards foreigners do not change with Erasmus. This suggests that the sense of group identity — national

versus European — does not significantly shift with the program. Still, the Erasmus program is highly selective and the effects presented in this article should be understood as the average treatment effects on the treated. Students in my sample who do not plan to participate in the study abroad program show a significant bias against foreigners in the Triple Dictator Game and they feel less "European" than students who are ready to go abroad. There could be potential for the program to increase a sense of European identity, if it were able to target students who feel less "international" to begin with.

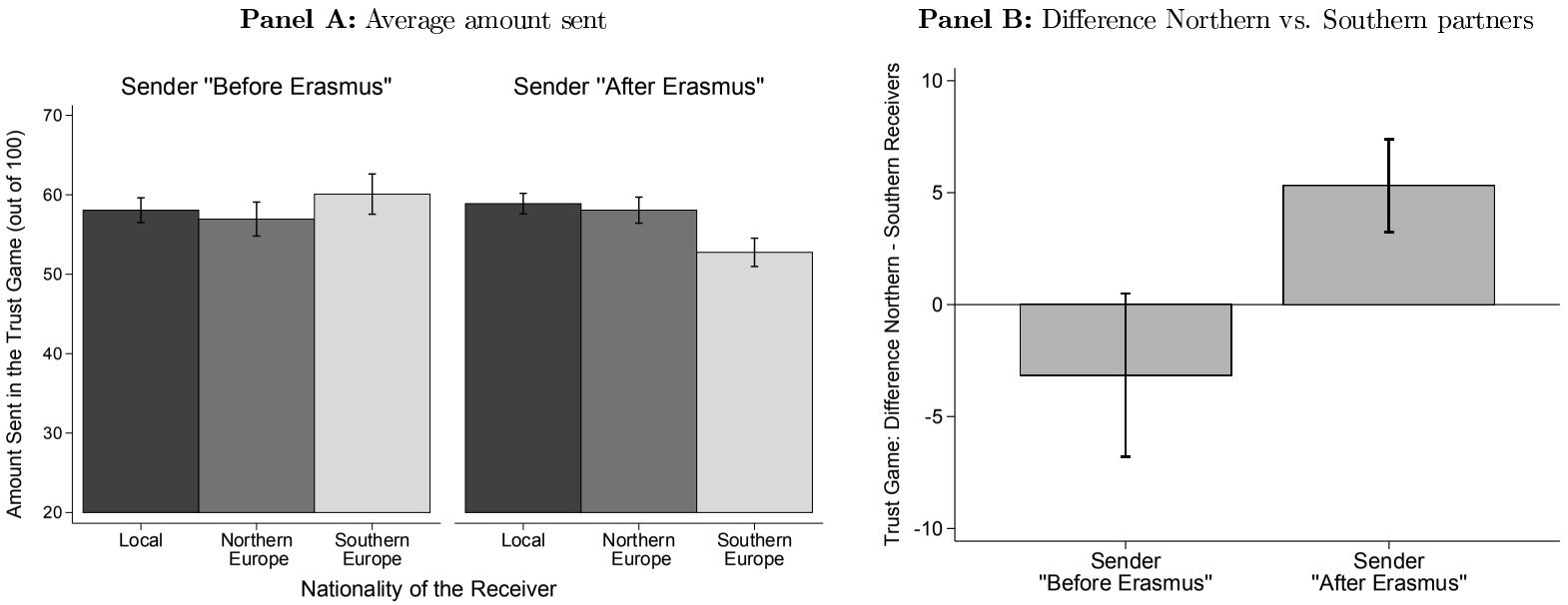
Overall, this paper confirms that individual attitudes towards people from other groups — nationalities in this case — can change simply by increased exposure to these groups. However, contrary to most studies on inter-group contact (Allport 1954; Pettigrew 1998; Pettigrew and Tropp 2006), I find that higher exposure leads to more discrimination. This seems to be driven by an increase in statistical discrimination, a channel which is not typically taken into account.

Figure 1.1: Differences in interpersonal trust across Europe



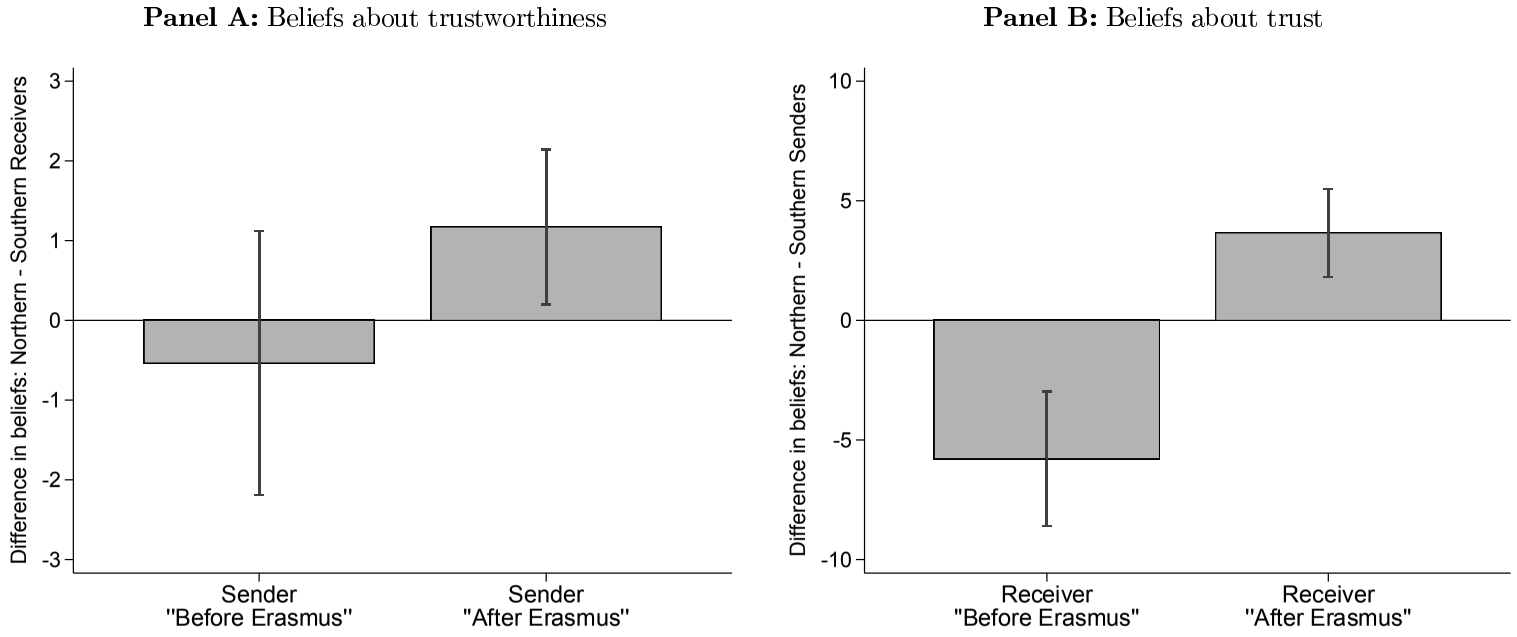
Notes: The Figure summarizes answers to the World Values Survey question "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?" (Data Source: ASEP/JDS). Bars indicate mean ± standard error.

Figure 1.2: Trust Game: Amount sent to partners of different nationalities, students before and after an Erasmus study abroad



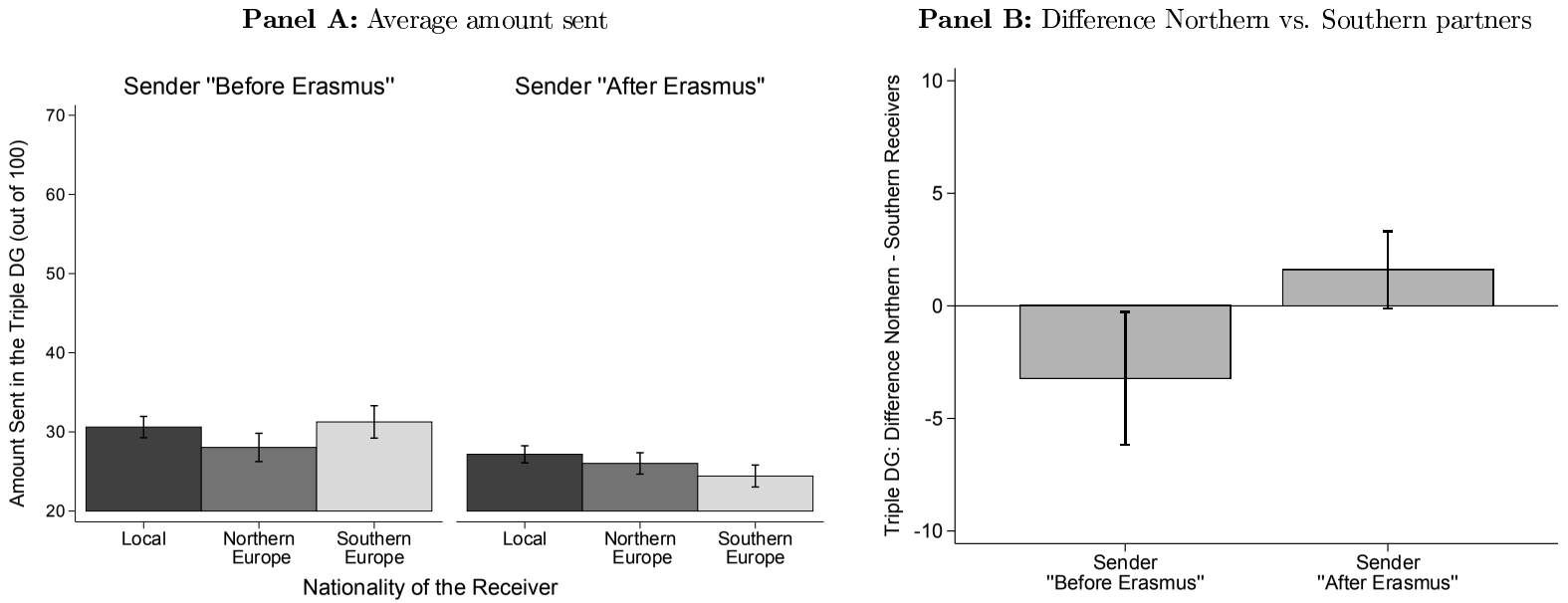
Notes: Panel A presents average amounts sent to partners of different nationalities in the Trust Game, disentangled by whether the subjects are about to leave on a study abroad stay (Sender "Before Erasmus") or have just returned from a study abroad stay (Sender "After Erasmus"). See Table 1.2 for the classification of countries into "Local", "North" and "South". Panel B presents differences in behavior towards Nothern vs. Southern partners. Bars indicate mean ± standard error.

Figure 1.3: Effect of an Erasmus study abroad stay on beliefs about trustworthiness and trust of Northern versus Southern partners



Notes: Panel (a) captures senders' beliefs about the amount returned by Northern vs. Southern receivers in the Trust Game (as % of what the partner received from the sender), disentangled by whether the subjects are about to leave on a study abroad stay (Sender "Before Erasmus") or have just returned from a study abroad stay (Sender "After Erasmus"). Panel (b) summarizes beliefs about the number of points received from Northern vs. Southern senders in the Trust Game, disentangled by the Erasmus status of the receiver. See Table 1.2 for the classification of countries into "North" and "South". Bars indicate mean ± standard error.

Figure 1.4: Triple Dictator Game: Amount sent to partners of different nationalities, students before and after an Erasmus study abroad



Notes: Panel A presents average amounts sent to partners of different nationalities in the Triple Dictator Game, disentangled by whether the subjects are about to leave on a study abroad stay (Sender "Before Erasmus") or have just returned from a study abroad stay (Sender "After Erasmus"). See Table 1.2 for the classification of countries into "Local", "North" and "South". Panel B presents differences in behavior towards Nothern vs. Southern partners. Bars indicate mean ± standard error.

Table 1.1: Sample characteristics

			"Before Erasmus" Mean (1)	"Before Erasmus" SD (2)	"After Erasmus" Mean (3)	"After Erasmus" SD (4)	T-test Difference (5)	T-test p-value (6)
Gender	Female	%	61,3	49,0	56,5	49,8	4,9	0,501
Age			22,6	1,5	23,9	1,6	-1,3***	0,000
Field of study	Business, Economics or Law	%	17,3	38,1	21,8	41,4	-4,4	0,451
	Humanities, Social Sciences or Education	%	32,0	47,0	30,7	46,3	1,4	0,843
	Math, Physics, Natural Sciences or Technical	%	16,0	36,9	21,0	40,9	-5,0	0,390
	Medicine	%	14,7	35,6	9,7	29,7	5	0,288
	Arts, Philosophy or Languages	%	20,0	40,3	16,9	37,7	3,1	0,589
Host Country	North	%	65,3	47,9	66,1	47,5	-0,8	0,909
	South	%	29,3	45,5	25,8	43,9	3,5	0,590
	New EU	%	5,3	22,6	8,1	27,3	-2,7	0,468
Participants	Total	N	75		124		199	

Notes: Means. The table presents characteristics of (Czech and Slovak) subjects before Erasmus study abroad program ("Before Erasmus", Column 1-2) and After Erasmus study abroad ("After Erasmus", Column 3-4).

Table 1.2: Classification of countries used in the analysis

"Local"	**"Foreign"**	
	"North"	**"South"**
Czech Rep.	Austria	France
Slovakia	Belgium	Greece
	UK	Italy
	Netherlands	Portugal
	Germany	Spain
	Ireland	

Notes: Only countries of origin for at least one participant in the experiment are listed.

Table 1.3: Means, across experimental manipulations

	(1)	(2)	(3)
Panel A: Amount sent in the Trust Game			
Sample	Senders "Before Erasmus"	Senders "After Erasmus"	P.p. Diff (2)-(1) (p-value)
Receiver Local	58.1	58.9	0.8 (0.677)
Receiver Foreign	58.3	55.6	-2.7 (0.197)
Receiver North	56.9	58.1	1.1 (0.652)
Receiver South	60.1	52.8	-7.3 (0.019)
P.p. diff Local-Foreign (p-value)	-0.2 (0.977)	3.3 (0.056)	
N	1,111	1,834	
P.p. diff North-South (p-value)	-3.1 (0.321)	5.3 (0.029)	
N	523	989	
Panel B: Amount sent in the Triple Dictator Game			
Sample	Senders "Before Erasmus"	Senders "After Erasmus"	P.p. Diff (2)-(1) (p-value)
Receiver Local	30.6	27.2	-3.4 (0.075)
Receiver Foreign	29.4	25.3	-4.1 (0.004)
Receiver North	28.0	26.0	-2.0 (0.285)
Receiver South	31.3	24.4	-6.8 (0.002)
P.p. diff Local-Foreign (p-value)	1.2 (0.829)	1.9 (0.154)	
N	1,111	1,834	
P.p. diff North-South (p-value)	-3.2 (0.131)	1.6 (0.506)	
N	523	989	

Notes: Means. Panel A reports amounts sent in the Trust Game, while Panel B reports amounts sent in the Triple Dictator game. The Table presents behavior of senders (Czech and Slovak) before ("Before Erasmus", column 1) and after Erasmus study abroad ("After Erasmus", column 2), by the nationality of the receiver. See Table 1.2 for the classification of countries into "North", "South", "Local" and "Foreign". All differences are presented in percentage points and tested using a Wilcoxon rank-sum test.

Table 1.4: Trust Game — Effects of Erasmus study abroad on behavior towards partners from Northern and Southern Europe

Dependent variable	Amount sent in the Trust Game					
Sample	"Before Erasmus" and "After Erasmus"					
	(1)	(2)	(3)	(4)	(5)	(6)
After Erasmus	1.12	-1.77	-1.87	2.05	0.98	0.91
	(5.36)	(5.56)	(5.55)	(4.78)	(5.07)	(5.12)
Receiver South	3.15	3.57	3.18	1.90	2.51	1.91
	(3.62)	(3.45)	(3.37)	(3.16)	(3.10)	(2.97)
Receiver South*After Erasmus	-8.46**	-9.11**	-8.54**	-6.53*	-7.07*	-6.22*
	(4.18)	(4.10)	(4.07)	(3.75)	(3.77)	(3.70)
Amount sent in the Triple DG				0.43***	0.42***	0.43***
				(0.05)	(0.06)	(0.06)
Constant	56.94***			44.86***		
	(4.34)			(4.37)		
Sender's gender, age, study major		yes	yes		yes	yes
Receiver's gender, age, study major		yes	yes		yes	yes
Order of the games, roles			yes			yes
Observations	1,512	1,512	1,512	1,512	1,512	1,512

Notes: OLS, standard errors are clustered on the sender level.* p<0.10, ** p<0.05, *** p<0.01. The estimation sample is comprised of (Czech and Slovak) subjects before and after Erasmus study abroad. The dummy variable "After Erasmus" is equal to one for subjects after and zero for those before the Erasmus program. The dummy variable "Receiver South" is equal to one if the receiver comes from Southern Europe and zero for receivers from Northern Europe. See Table 1.2 for the classification of countries into "North" and "South". In Columns 1-6, the omitted group are decisions of subjects "Before Erasmus" towards receivers from Northern Europe.

Table 1.5: Trust Game — Effects of Erasmus study abroad, by the region of the Erasmus stay

Dependent variable	Amount sent in the Trust game						
Sample	All	Stay North			Stay South		
VARIABLES	Senders "Before" and "After Erasmus" (1)	Senders "Before" and "After Erasmus" (2)	Senders "Before Erasmus" (3)	Senders "After Erasmus" (4)	Senders "Before" and "After Erasmus" (5)	Senders "Before Erasmus" (6)	Senders "After Erasmus" (7)
After Erasmus	1.12	-4.15			9.77		
	(5.36)	(6.68)			(9.59)		
Receiver South	3.15	0.04	0.04	-5.64**	11.64*	11.64	-3.45
	(3.62)	(4.45)	(4.48)	(2.57)	(6.86)	(6.95)	(4.43)
After Erasmus*Receiver South	-8.46**	-5.68			-15.09*		
	(4.18)	(5.14)			(8.15)		
Constant	56.94***	61.59***	61.59***	57.44***	46.29***	46.29***	56.06***
	(4.34)	(5.39)	(5.42)	(3.96)	(7.46)	(7.56)	(6.06)
Observations	1,512	1,010	348	662	403	147	256

Notes: OLS, standard errors are clustered on the sender level.* p<0.10, ** p<0.05, *** p<0.01. The estimation sample is comprised of (Czech and Slovak) subjects before and after Erasmus study abroad. Columns 2-4 analyse behavior of subjects with planned or completed stay in Northern Europe, while Columns 5-7 focus on subjects going to Southern Europe. The dummy variable "After Erasmus" is equal to one for subjects after and zero for those before the Erasmus program. The dummy variable "Receiver South" is equal to one if the receiver comes from Southern Europe and zero for receivers from Northern Europe. See Table 1.2 for the classification of countries into "North" and "South". In Columns 1,2,5, the omitted group are decisions of subjects "Before Erasmus" towards receivers from Northern Europe. In Columns 3-4 and 6-7, the omitted group are decisions towards receivers from Northern Europe.

Table 1.6: Triple Dictator Game — Effects of Erasmus study abroad on behavior towards partners from Northern and Southern Europe

Dependent variable	Amount sent in the Triple DG		
Sample	"Before Erasmus" and "After Erasmus"		
	(1)	(2)	(3)
After Erasmus	-2.02	-6.39	-6.25
	(4.44)	(4.47)	(4.40)
Receiver South	3.22	3.07	3.47
	(2.94)	(2.65)	(2.47)
Receiver South*After Erasmus	-4.82	-5.32*	-5.78*
	(3.40)	(3.08)	(2.99)
Constant	28.04***		
	(3.51)		
Sender's gender, age, study major		yes	yes
Receiver's gender, age, study major		yes	yes
Order of the games, roles			yes
Observations	1,512	1,512	1,512

Notes: OLS, standard errors are clustered on the sender level.* p<0.10, ** p<0.05, *** p<0.01. The estimation sample is comprised of (Czech and Slovak) subjects before and after Erasmus study abroad. The dummy variable "After Erasmus" is equal to one for subjects after and zero for those before the Erasmus program. The dummy variable "Receiver South" is equal to one if the receiver comes from Southern Europe and zero for receivers from Northern Europe. See Table 1.2 for the classification of countries into "North" and "South". In Columns 1-3, the omitted group are decisions of subjects "Before Erasmus" towards receivers from Northern Europe.

Table 1.7: Outgroup bias against foreigners — Effects of an Erasmus study abroad

Dependent variable	Amount sent in the Triple DG					
Sample	Senders "Before Erasmus" and "After Erasmus"					
	(1)	(2)	(3)	(4)	(5)	(6)
After Erasmus	-3.44	-8.80**	-8.99**	-3.44	-8.81**	-9.01**
	(4.31)	(4.17)	(4.08)	(4.31)	(4.17)	(4.08)
Receiver Foreign	-1.20	-1.75	-2.17			
	(1.40)	(1.38)	(1.37)			
Receiver Foreign*After Erasmus	-0.69	0.00	0.18			
	(1.88)	(1.81)	(1.81)			
Receiver North				-2.57	-2.99*	-3.68**
				(1.70)	(1.66)	(1.52)
Receiver South				0.65	-0.07	-0.12
				(2.37)	(2.16)	(2.12)
Receiver North*After Erasmus				1.42	2.42	2.85
				(2.19)	(2.09)	(2.03)
Receiver South*After Erasmus				-3.40	-3.06	-3.20
				(2.90)	(2.67)	(2.65)
Constant	30.61***			30.61***		
	(3.37)			(3.37)		
Sender's gender, age, study major		yes	yes		yes	yes
Receiver's gender, age, study major		yes	yes		yes	yes
Order of the games, roles			yes			yes
Observations	2,945	2,945	2,945	2,945	2,945	2,945

Notes: OLS, standard errors are clustered on the sender level.* p<0.10, ** p<0.05, *** p<0.01. The estimation sample is comprised of (Czech and Slovak) subjects before and after Erasmus study abroad. The dummy variable "After Erasmus" is equal to one for subjects after and zero for those before the Erasmus program. The dummy variable "Receiver Foreign" is equal to one if the receiver comes from abroad and zero for local receivers. See Table 1.2 for the classification of countries into "Local", "Foreign", "North" and "South". In columns 1-6, the omitted group are decisions of subjects "Before Erasmus" towards "Local" receivers.

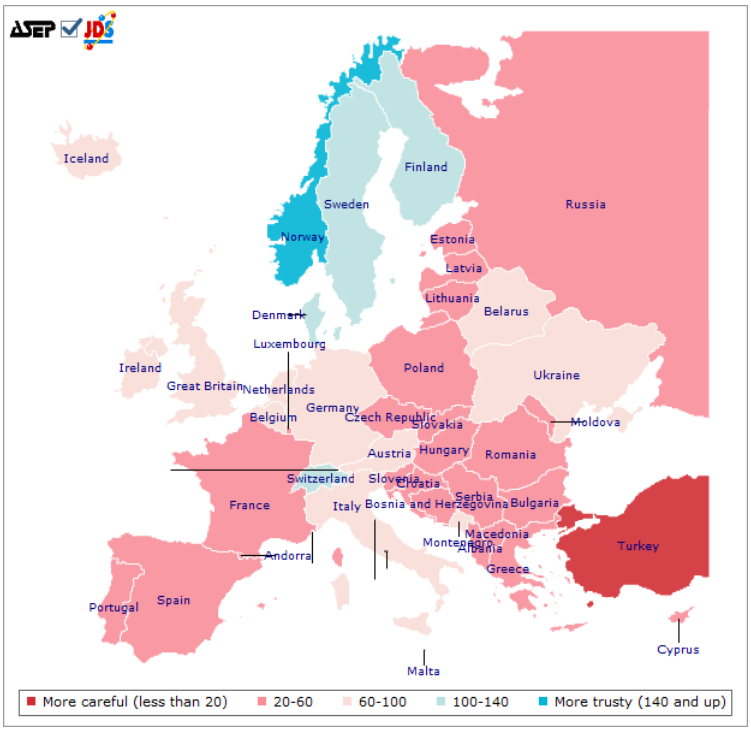
Table 1.8: Outgroup bias against foreigners — Effects of an Erasmus study abroad vs. selection effect

Dependent variable	Amount sent in the Triple DG		
Sample	Senders "Before Erasmus"	Senders "After Erasmus"	Senders "Never Erasmus"
	(1)	(2)	(3)
Receiver Foreign	-1.67	-1.84	-3.74**
	(1.32)	(1.31)	(1.81)
Sender's gender, age, study major	yes	yes	yes
Receiver's gender, age, study major	yes	yes	yes
Order of the games, roles	yes	yes	yes
Observations	1,111	1,834	781

Notes: OLS, standard errors are clustered on the sender level.* p<0.10, ** p<0.05, *** p<0.01. The estimation sample is comprised of all local (Czech and Slovak) subjects — before Erasmus study abroad program ("Before Erasmus", column 1), After Erasmus study abroad ("After Erasmus", column 2) and program non-participants ("Never Erasmus", column 3). The dummy variable "Receiver Foreign" is equal to one if the receiver comes from abroad and zero for local receivers. See Table 1.2 for the classification of countries into "Local" and "Foreign". In Columns 1-3, the omitted group are decisions towards "Local" receivers.

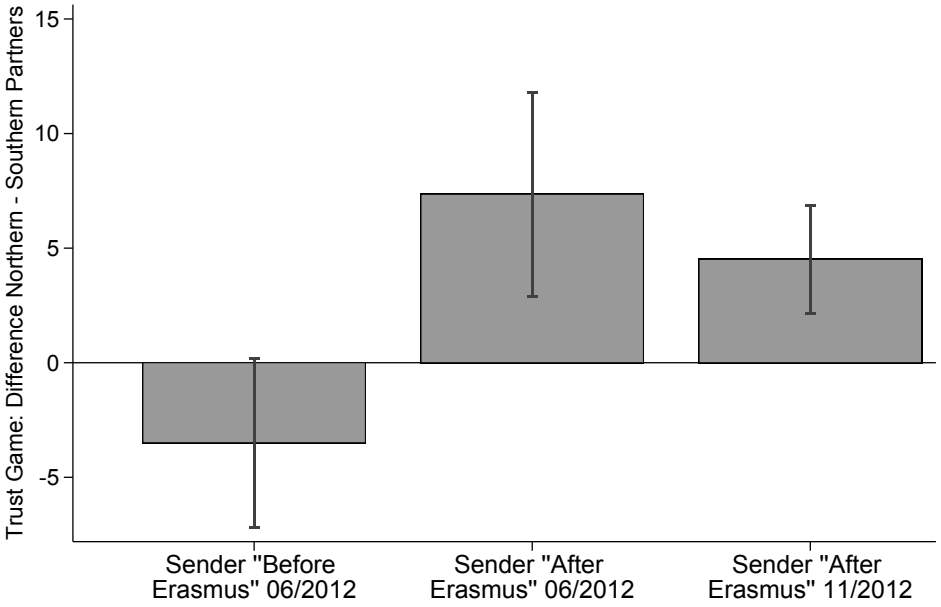
1.A Appendix 1

Figure 1.5: Interpersonal trust across European countries



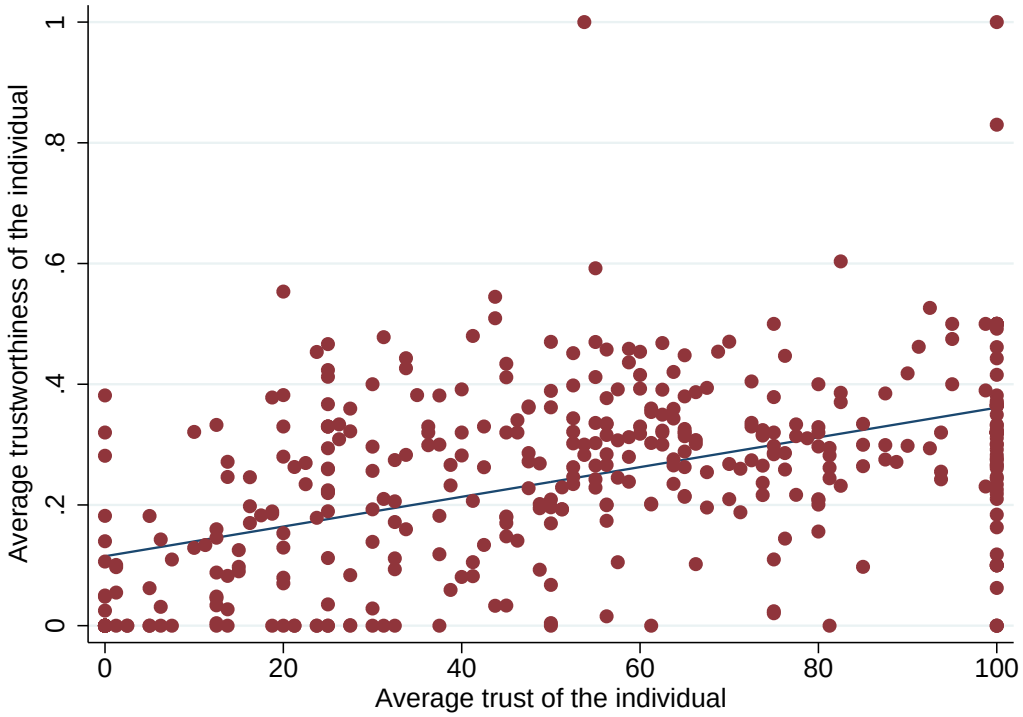
Notes: The Figure summarizes answers to the World Values Survey question "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?" (Source: ASEP/JDS)

Figure 1.6: Robustness check — Effects of Erasmus study abroad on behavior towards partners from Northern and Southern Europe, by the date of the experiment



Notes: The Figure presents differences in behavior towards Northern vs. Southern partners in the Trust Game, disentangled by whether the subjects are about to leave on a study abroad stay (Sender "Before Erasmus") or have just returned from a study abroad stay (Sender "After Erasmus") and by the time of the experiment (June 2012 vs. November 2012). See Table 1.2 for the classification of countries into "North" and "South". Bars indicate mean ± standard error.

Figure 1.7: Correlation between individual trust and trustworthiness for all subjects in the experiment



Notes: Average individual trust is calculated as the average amount sent in the Trust Game, averaging over the 16 profiles of potential partners. Average individual trustworthiness is calculated as average return ratio (*Return ratio=amount returned to sender/(3*amount sent by sender)*, averaging over all receiver's decisions. Each receiver made 80 trustworthiness decisions — there are 16 profiles of potential senders and 5 trustworthiness decisions per sender, as receivers' decisions were elicited using a strategy method.

Table 1.9: Recruitment into the experiment — all Charles University outbound Erasmus students vs. experiment participants

			Database of outbound Erasmus students		Experiment participants		Difference databases	Recruitment into the experiment	
			2012/13	2011/12	2012/13	2011/12	diff (2)-(1)	diff(3)-(1)	diff (4)-(2)
			(1)	(2)	(3)	(4)	(5)	(6)	(7)
Gender	Female	%	69,7	69,9	61,3	56,8	0,2	8,3	13,1***
Study program	BA	%	38,1	28,1	49,3	33,9	-10,0***	-11,2*	-5,8
	MA	%	59,0	69,0	49,3	65,3	10,0***	9,7	3,7
Field of study	Business/Economics/Law	%	14,8	17,2	17,3	22,9	2,4	-2,5	-5,7
	Humanities/Social sc./Education	%	31,6	30,4	32,0	32,2	-1,2	-0,4	-1,8
	Math/Physics/Natural sc./Technical	%	15,6	15,0	16,0	18,6	-0,6	-0,4	-3,6
	Medicine	%	16,0	18,8	14,7	10,2	2,8	1,3	8,6**
	Arts/Philosophy/Languages	%	21,9	18,5	20,0	16,1	-3,4	1,9	2,4
Host Country	North	%	64,4	60,9	65,3	66,1	-3,5	-0,9	-5,2
	South	%	30,2	32,7	29,3	26,27	2,5	0,9	6,4
	New EU	%	5,4	6,4	5,3	7,63	1,0	0,1	-1,2
Participants	Total	N	923	1009	75	118	1932	998	1127

Notes: Means. "Before Erasmus" subjects were recruited from the 2012/13 database, "After Erasmus" subjects were recruited from the 2011/12 database. Six students from the "After Erasmus" sample were not students of Charles University or they participated in the Erasmus program prior to the academic year 2011/2012; for these reasons they are not included in this comparison. All differences are tested using a t-test, *p<0.10, ** p<0.05, *** p<0.01.

Table 1.10: Trust Game — Effects of Erasmus study abroad on behavior towards partners from Northern and Southern Europe, ordered probit

	Marginal fixed effects after ordered probit					
Dependent variable	Probability of the Amount sent in the Trust Game being:					
	100	80	60	40	20	0
Sample	Senders "Before Erasmus" and "After Erasmus"					
	(1)	(2)	(3)	(4)	(5)	(6)
After Erasmus	0.010	0.001	0.000	-0.001	-0.002	-0.008
	(0.053)	(0.006)	(0.000)	(0.007)	(0.011)	(0.042)
Receiver South	0.030	0.003	-0.000	-0.004	-0.006	-0.023
	(0.036)	(0.004)	(0.001)	(0.005)	(0.008)	(0.028)
Receiver South*After Erasmus	-0.080**	-0.011*	-0.002	0.010**	0.016**	0.066*
	(0.040)	(0.006)	(0.003)	(0.005)	(0.008)	(0.035)
Observations	1,512	1,512	1,512	1,512	1,512	1,512

Notes: Ordered probit, standard errors are clustered on the sender level.* p<0.10, ** p<0.05, *** p<0.01. The estimation sample is comprised of (Czech and Slovak) subjects before and after Erasmus study abroad. The dummy variable "After Erasmus" is equal to one for subjects after and zero for those before the Erasmus program. The dummy variable "Receiver South" is equal to one if the receiver comes from Southern Europe and zero for receivers from Northern Europe. See Table 1.2 for the classification of countries into "North" and "South". In Columns 1-6, the omitted group are decisions of subjects "Before Erasmus" towards receivers from Northern Europe.

Table 1.11: Beliefs about trust — Effects of Erasmus study abroad on beliefs about behavior of Senders from Northern and Southern Europe

Dependent variable	Beliefs about Amount sent in the Trust Game				
Sample	Receiver "Before Erasmus" and "After Erasmus"	Receiver "Before Erasmus"	Receiver "After Erasmus"	Receiver "Before Erasmus" and "After Erasmus"	
	(1)	(2)	(3)	(4)	(5)
Receiver After Erasmus	1.54			-1.10	-1.45
	(4.73)			(4.84)	(4.87)
Sender South	5.78**	5.78**	-3.65**	6.01**	5.43**
	(2.81)	(2.82)	(1.84)	(2.70)	(2.63)
Sender South*Receiver After Erasmus	-9.43***			-9.88***	-9.36***
	(3.35)			(3.33)	(3.31)
Constant	54.22***	54.22***	55.76***		
	(3.83)	(3.85)	(2.78)		
Sender's gender, age, study major				yes	yes
Receiver's gender, age, study major				yes	yes
Order of the games, roles					yes
Observations	1,512	523	989	1,512	1,512

Notes: OLS, Standard errors are clustered on the receiver level.* p<0.10, ** p<0.05, *** p<0.01. Beliefs about trust capture beliefs of receivers regarding the amount sent to them by the sender in the Trust Game. The estimation sample is comprised of (Czech and Slovak) subjects before and after Erasmus study abroad. The dummy variable "Receiver After Erasmus" is equal to one for subjects after and zero for those before the Erasmus program. The dummy variable "Sender South" is equal to one if the sender comes from Southern Europe and zero for the senders from Northern Europe. See Table 1.2 for classification of countries into "North" and "South". In columns 1,4,5, the omitted group are beliefs of subjects "Before Erasmus" regarding behavior of senders from Northern Europe. In columns 2-3, the omitted group are beliefs regarding senders from Northern Europe.

Table 1.12: Trust Game — Behavior of foreign subjects from Northern and Southern Europe

Dependent variable	Amount sent in the Trust Game					
Sample	Foreigners from Northern Europe			Foreigners from Southern Europe		
	(1)	(2)	(3)	(4)	(5)	(6)
Receiver North	-0.06	0.72	0.85	-1.19	-1.26	-1.33
	(2.76)	(2.54)	(2.55)	(2.17)	(2.50)	(2.42)
Receiver South	-11.15***	-9.80***	-9.55***	1.40	0.76	0.91
	(3.16)	(3.61)	(3.44)	(3.22)	(3.48)	(3.42)
Constant	57.98***			48.31***		
	(4.84)			(3.75)		
Sender's gender, age, study major		yes	yes		yes	yes
Receiver's gender, age, study major		yes	yes		yes	yes
Order of the games, roles			yes			yes
Observations	671	671	671	1,155	1,155	1,155

Notes: OLS, standard errors are clustered on the sender level.* p<0.10, ** p<0.05, *** p<0.01. The estimation sample is comprised of foreign subjects from Northern Europe in Columns 1-3 and foreign subjects from Southern Europe in Columns 4-6. The dummy variable "Receiver North" is equal to one if the receiver comes from Northern Europe and zero otherwise and variable "Receiver South" is equal to one if the receiver comes from Southern Europe and zero otherwise. In Columns 1-6, the omitted group are decisions of foreign subjects towards "Local" receivers. See Table 1.2 for the classification of countries into "Local", "North" and "South".

Chapter 2

Social Contagion of Ethnic Hostility

Co-authored by Michal Bauer, Julie Chytilová and Tomáš Želinský

2.1 Introduction

Intergroup conflict is one of the most pressing problems facing the world, giving rise to phenomena such as civil wars, ethnic cleansing, and discrimination. History of the Holocaust, the wars in Rwanda, the Balkans, and the Middle East, for example, underscore the importance of large-scale inter-group hostility. Ethnic animosities often spread quickly (Glaeser 2005; Gentzkow and Shapiro 2004). A major puzzle about the escalation of ethnic conflict is why it is relatively common to see people suddenly changing behavior from cooperating with people across ethnic lines to taking an active part in ethnic aggression (Esteban and Ray 2008; Basu 2005; Yanagizawa-Drott 2014; Fearon and Laitin 2000). For instance, Bardhan (2005, p.169) describes the issue as follows: "[It] is not uncommon to see communities sharing some historical animosities coexisting peacefully [...] for generations (Serbs, Croats and Muslims in the former Yugoslavia, for example) and then something snaps and inter-community violence erupts." This makes it essential to understand the role of short-lived changes in economic and social environment in shaping individual willingness to engage in ethnic hostility.

In this paper, we explore how an individual decision whether to do harm to others is influenced by the actions of peers within their own social group. The social aspect is a ubiquitous feature of situations when ethnic hostilities are manifested: bullying in schools,

harassment of ethnic minorities or violent attacks during wars are typically performed side by side with others from their own social group. The main question of this paper is the following: Is ethnic hostility contagious? Specifically, we explore whether (i) susceptibility to follow peers is intensified when choices impact a dissimilar ethnic group as compared to a co-ethnic group, and if such asymmetry exists, (ii) whether such elevated conformism is specific for destructive, conflict-like interactions. To answer these questions, we study discrimination against Roma, the largest ethnic minority in Europe, among adolescents from a majority population. We use incentivized tasks (Joy of Destruction game and Prisoner's Dilemma game) to identify individual willingness to be hostile and to cooperate. A novel feature of our design, relative to other experiments measuring discrimination,[1] is that we randomly match real-life peers, let them make choices sequentially and thus a large fraction of the subjects observes the peer's choice prior to making their own decision. This allows us to identify not only a baseline level of discrimination but also differences in susceptibility to follow peers. The second distinguishing feature is our focus on discrimination in hostile behavior, i.e. whether to cause harm.

On the conceptual side, our paper aims to bring together two important literatures on how the social environment may affect individual choice: one focusing on social motivations behind peer effects and the other one highlighting the link between intergroup-conflict (or the threat of it) and parochial altruism. Theory, as well as evidence, suggests that the social aspects of individual choice become more important in the context of inter-group conflict. Evolutionary theories have long emphasized the role of inter-group conflict in the development of parochial altruism — willingness to cooperate and suppress conflict with in-group members combined with a preference for favoring the members of one's ethnic or racial group (Choi and Bowles 2007; Bowles 2008; Bernhard, Fischbacher, and Fehr 2006). Since individual survival in inter-group conflicts is often linked to the fate of his group, these approaches suggest that people may be sensitive to signals or direct experiences of conflict between groups, and such experiences may increase in-group cohesion by increasing altruism, norm adherence or the punishment of norm violators.

[1] Audit studies and correspondence tests help to uncover the existence of ethnic discrimination on various markets (Yinger 1998; Bertrand and Mullainathan 2004; Bartoš et al. 2015), but do not allow researchers to manipulate the social context of the choice. The existing lab and artefactual field experiments made important progress in studying discrimination against people with different group attributes in a controlled environment, but so far focused on studying choices of individuals made in isolation from others (Fershtman and Gneezy 2001; Bernhard, Fischbacher, and Fehr 2006; Berge et al. 2015a; Falk and Zehnder 2013; Goette, Huffman, and Meier 2006; Angerer et al. 2015; Cahlíková 2015), and thus abstracting from social influences.

On the empirical front, a recent body of detailed micro-level studies from a range of post-conflict societies demonstrates a strong pattern: greater individual exposure to war violence tends to increase cooperative behavior at the local level, such as local community participation (Bellows and Miguel 2009), voting in elections (Blattman 2009), and greater concern about fairness in behavioral tasks with in-group members (Voors et al. 2012; Bauer et al. 2014; Gilligan, Pasquale, and Samii 2014), suggesting that exposure to conflict increases group cohesion.[2]

At the same time, a leading explanation behind conformism, a tendency to make the same choices as those made by peers, is based on the idea that there are utility gains from following others due to either social preferences (gaining utility from pleasing others) and social pressure (avoiding future punishment or building a reputation of being supportive), precisely the components which group conflicts are predicted to intensify. In this paper, we hypothesize that elevated group cohesion can be triggered not only by the objective existence or past experience of inter-group conflict (as documented by previous work), but also by more subtle signals of threat of conflict, such as ethnically-motivated harmful actions of peers, and that the behavioral response, increased conformism, can contribute to the spreading of hostility.

The issue is of fundamental importance for a policy aiming to prevent social unrest and conflicts. If doing harm to an ethnic minority is particularly contagious, early diagnosis and intervention seems crucial to prevent the spreading of such acts. Moreover, many countries have adopted laws which impose additional punishment for racially-motivated crimes or hate speech. Besides the fairness concerns, the justification of these policies rests on the assumption that such acts have wider impacts on society beyond the immediate victim by making hostilities easier to spread and ultimately threaten social cohesion. It is therefore important to know whether such negative externality exists in practice. Yet, little is known about the influence of others when deciding whether to engage in aggressive behavior towards ethnic out-groups empirically. This is mainly because it requires a researcher not only to causally identify the influence of destructive actions on the actions of other people, but also to identify whether the spreading of destructive actions is amplified when a victim is from an ethnic group other than the perpetrator, a task that is extremely difficult to achieve using naturally occurring data. Although destruc-

[2]Note that the theory of Choi and Bowles (2007) and Bowles (2008) postulates that the development of parochial altruism due to inter-group conflict operates at the evolutionary level, i.e. due to selection. Here, as in the above-cited papers, we consider an individual behavioral reaction of being exposed to conflict in the spirit of these models.

tive behavior in our experiment is more innocuous than hate crime, our experimental design contains three key elements: (i) variation in ethnicity of the victim, (ii) random assignment to observing hostile actions of peers, and (iii) an opportunity to follow and act destructively too.

We elicited hostile behavior in incentivized tasks conducted among adolescents (N=327) from a majority ethnic group in Eastern Slovakia (artefactual field experiments), a region characterized by a high risk of escalation of anti-Roma protests or violence. The situation of the Roma, a minority of Indian origin and estimated size of 10-12 million in the EU, is considered one of the most pressing social and human rights issues in the EU (European Commission 2004). We sampled subjects 13–15 years old. Understanding the behavior of this age group is important because, as Levin and McDevitt (2002, p. 196) describe: "The first characteristic that distinguishes hate crime offenders from other serious offenders is their age. Most research to date indicates that the vast majority of perpetrators are very young males, often juveniles."[3]

To measure individual willingness to engage in hostile behavior, we administered the Joy of Destruction game: a money-burning experiment (Abbink and Herrmann 2011; Abbink and Sadrieh 2009; Prediger, Vollan, and Herrmann 2014), in which two players receive the same endowment and simultaneously choose whether to pay to reduce the other player's payoff below one's own. The task is tailor-made to identify hostile behavior, since neither selfish nor fairness motives can justify destruction. Since subjects make decisions without knowing what the other player does, beliefs about destructive behavior of the other player may motivate pre-emptive hostility. In addition to this task, we implemented a Prisoner's dilemma game, in order to measure willingness to cooperate.

To identify ethnic discrimination, we manipulated whether an individual's (from the majority group) decision impacts an unknown co-ethnic or an unknown member of dissimilar ethnic minority (Roma), using names to signal ethnicity.[4] At the same time, we exogenously vary the social context of individual choice. Decisions were made either in isolation from peers or in groups of three individuals who made choices sequentially and could observe each other. The experiments were implemented in a natural environment for this subject pool — schools, which allowed us to match individuals with real-life peers

[3]In 2013, 20% of hate crime offenders in Slovakia were between 15–18 years old, in the Czech Republic 26% were 15–20 years old, and in US 32% were under 18 years old. In Canada, 38% of those accused of hate crime in 2006 were 12–17 years old and in Sweden, 40% of those suspected of involvement in hate crimes were younger than 20 years old.

[4]Notice that we do not measure discrimination based on personal knowledge of the partner, since subjects are always matched with a partner from an unknown distanced school.

who live in similar socio-economic conditions. Since participants could not choose with whom they would be matched and in what order they make decisions, the design allows estimating the causal effect of peer behavior on individual actions.

The main results can be summarized as follows. First, we find that subjects do not discriminate against the ethnic minority when making choices in isolation: the prevalence of destructive choices is non-negligible but it is unaffected by ethnicity of the counterpart. Second, peers are very influential and the susceptibility to follow is particularly strong when a hostile action harms an ethnic minority — being exposed to a peer who acts aggressively instead of one who chooses not to destroy increases the likelihood of destructive choice by 60 percentage points. The proclivity to follow is also positive but less-than-half in magnitude when choice affects an unknown co-ethnic.

Third, we find that a greater influence of peers on willingness to engage in destructive behavior towards an ethnic minority, as compared to co-ethnics, is driven by a greater susceptibility to follow destructive behavior, and not by differences in susceptibility to follow non-destructive behavior. When choices impact the ethnic minority, the effect of observing a destructive peer (relative to not receiving any signal) increases the likelihood of destroying by 36 percentage points, while the effect is close to zero when choices affect a co-ethnic. As a consequence, ethnic discrimination in the willingness to cause harm arises when peers choose to harm, while we find virtually no discrimination when peers are peaceful. This result is based on a large sample of observations and it holds when an individual observes behavior of one, as well as two, peers. Interestingly, the main effect is somewhat larger for male decision-makers relative to females. Last, in the Prisoner's Dilemma game we find virtually no ethnic discrimination in cooperation and we also do not find disproportionate tendency to follow the behavior of others, suggesting that ethnically motivated herding is specific for a hostile type of interaction.

In terms of mechanisms, the full set of findings is consistent with the idea that harmful actions of peers against another ethnic group trigger a parochial psychological response and thus cement in-group conformism. We also discuss potential alternative explanations based on social learning and uncertainty about actions of members of the ethnic minority or uncertainty about social norms.

Our findings speak to several streams of literature. The paper contributes to the emerging literature, which aims to identify factors that shape the prevalence and intensity of ethnic discrimination. We find no discrimination in terms of cooperative as well as destructive behavior when subjects make a decision in isolation or when they observe

peaceful peers. These findings resonate with recent studies that find surprisingly little ethnic discrimination when individuals make choices in behavioral tasks in other settings. Perhaps most strikingly, Berge et al. (2015a) find virtually no evidence of co-ethnic bias in a dictator and public goods games among a large sample of working-class individuals in Kenya, a setting with recent history of ethnic violence. Also, Angerer et al. (2015) find only mild discrimination based on language among German and Italian-speaking children in Northern Italy.[5]

At the same time, our results are consistent with several pieces of evidence documenting that environmental factors associated with threat and conflict may be important drivers of ethnic discrimination. Hjort (2014) explores the efficiency of flower production in Kenya and finds a sharp increase in ethnic discrimination among workers during the period of ethnic conflict following the violent 2007 elections. Shayo and Zussman (2011) study the decisions of Jewish and Arab judges in Israel and find that ethnic bias in decisions increases with the intensity of terrorism in the vicinity of the court during the period preceding the ruling. Together, these diverse pieces of evidence are consistent with the idea that while relatively small latent ethnic biases may not be so important during everyday decisions, they can easily gain importance when combined with life experience or social influences which make a threat by other groups salient. Our results suggest that the unambiguously hostile actions of peers might also, similarly to violent elections or exposure to fighting, trigger feelings of threat, leading to the emergence of ethnic discrimination.

Our results are related to an important set of non-experimental studies which examine the individual willingness of civilians to do harm to other ethnic groups. In the context of Rwandan genocide, the existing work has documented the important role of political elites in mobilizing Hutu civilians to violence against Tutsi, via radio broadcasting (Yanagizawa-Drott 2014), public meetings (Bonnier et al. 2015), and pressure from governmental militia groups (Rogall 2015). DellaVigna et al. (2014) find the effects of Serbian radio broadcasting on voting for extreme nationalist parties in Croatia. Our paper provides evidence of the importance of the role of individuals within one's own social network in triggering engagement in hostile actions in behavioral tasks.

Last, our paper adds to the large literature on peer effects, which documented that

[5]Interestingly, Angerer et al. (2015) also find clear evidence of preferential treatment of individuals known to a decision-maker (in-group) relative to both types of partners that are unknown to the decision-maker (those with the same and distinct language background).

individual choices are often sensitive to what others are doing, with peers affecting student test scores (Kremer and Levy 2008; Sacerdote 2001), cooperative behavior (e.g. Gächter, Nosenzo, and Sefton 2013), and also socially harmful behavior, such as littering (Keizer, Lindenberg, and Steg 2008), cheating in exams (Gino, Ayal, and Ariely 2009), alcohol consumption, drug use and risky sexual behavior (Kremer and Levy 2008; Card and Giuliano 2013; Eisenberg, Golberstein, and Whitlock 2014), and participation in crime (Damm and Dustmann 2014; Bayer, Hjalmarsson, and Pozen 2009). A novelty of this paper is that we look at how the identity of the victim affects the strength of peer effects, i.e. whether conformity is greater when a victim is from an ethnic minority relative to a co-ethnic.

The paper proceeds as follows. In Section II we provide a brief background about Roma and the setting of our study. In Section III we describe the sample and experimental design. Section IV presents the findings. Section V provides a discussion about alternative interpretations and Section VI offers concluding remarks.

2.2 Background about Roma in Europe and Eastern Slovakia

Roma people (sometimes also referred to as Gipsy people) constitute the largest ethnic minority in Europe, estimated at 10–12 million. It is a minority that lives in poor socio-economic conditions and social exclusion all over Europe.[6] The average education levels of Roma population are low (15% finish upper-secondary education), they are poorly integrated into the labor market (less than one third are in paid employment), live in substandard housing, have worse health and lower life expectancy compared to the majority population. It is estimated that 90% of Roma in Europe live below the national poverty lines. Overall, the situation of Roma is considered one of the most pressing social and human rights issues in the EU (European Commission 2004).

In Eastern Slovakia, the setting we study, the concentration of Roma is high: estimated at 15% of the local population. Around 70% of Roma live segregated from the majority population, often in isolated settlements or on the edge of villages and towns. Over 90% of the Roma in Slovakia are at risk of poverty and around 55% live without at least one of the four following basic amenities: indoor kitchen, indoor toilet, shower/bath

[6]The most significant populations are in the Eastern European countries (Bulgaria, Romania, Slovakia, Hungary and the Czech Republic), but also in France, Greece and the U.K.

and electricity (FRA and UNDP 2012). Schools are characterized by high levels of segregation of Roma children — 58% attend schools or classes where the majority of their classmates are also Roma and 76% of the majority children living near the Roma attend classes that are not ethnically mixed. Less than 20% of Roma finish upper-secondary education and less than 30% have paid employment, compared to around 90% and 60%, respectively, for the majority population living in similar areas as Roma.

Despite substantial efforts to improve the situation of Roma during the past 10–15 years, Roma are still subject to prejudice and face discrimination on markets (Bartoš et al. 2015). According to reports by the European Commission, almost one quarter of Europeans (38% of Slovaks) state that they would be uncomfortable with 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. Roma are also frequently associated with crime.[7] Populist politicians commonly use Anti-Roma rhetoric, especially before elections. For instance, in 2008 the Italian government, led by Silvio Berlusconi, announced a plan to fingerprint all 150,000 Roma, as a measure to "crackdown on crime". In Slovakia, the far-right "People's Party Our Slovakia", which, in its official campaign documents, refers to "desperate villages and towns suffering from crime and terror from Gipsy extremists", obtained 55% of the votes in the 2013 elections and its leader became the regional governor in central Slovakia. Thus, Eastern Slovakia represents an apt natural setting for studying factors which facilitate the spreading of fear and hostility against a segregated ethnic minority.

After World War II, when Roma were targeted by similar policies and persecution to Jews, there has not been any systematic violent conflict involving Roma. Nevertheless, the frequency of anti-Roma violence has been increasing in the last few years, especially in Central and Eastern Europe (Council of Europe 2012). During the last four years a series of anti-Roma protests were staged in dozens of towns and cities across the region. These protests, in which far-right extremists were often joined by the local population, commonly escalated into property damage or violence. In addition, walls separating the majority population from their Roma neighbors have been built in at least fourteen different cities in Slovakia since 2008, similarly to other countries such as Romania and Bulgaria.

[7]For instance, more than half of the people in Hungary believe that "crime is in the 'gipsy' blood" (FXB Center for Health and Human Rights at Harvard University 2014). The media tend to report on the minority mostly in the context of crime and social issues (Council of Europe 2012).

2.3 Experimental design

2.3.1 Sample selection

The experiment was conducted in Eastern Slovakia, during September and October 2013. Our sample are adolescents from the majority population.[8] We identified 13 schools in villages and small towns with a Roma neighborhood or settlement (within a maximum distance of five kilometers) and classes with predominantly majority students. They are geographically spread across the region (the map in Figure 2.1 shows the location of field sites).

We sampled from a population of adolescents aged 13–15 (grade 8 and 9). This age group is mature enough to understand the experimental tasks, typically has experience of interacting with the Roma minority, and is just entering the age characterized by high rates of participation in bullying and hate crime, as described above. At the same time, since schooling in Slovakia is obligatory until the age of 16 and there are few selective tracks available prior this age,[9] organizing the experiments among the last grades in primary schools helps us to avoid problems that could arise due to self-selection in the experiment. Importantly, it allows us to exploit the fact that subjects, when making decisions, could be naturally matched with their classmates, i.e. their real-life peers who live in a similar social environment. Our sample is 327 subjects.[10] Participation in the experiment was voluntary and the subjects could leave at any time. All students decided to complete the tasks.[11]

Sample characteristics are presented in Column 1 of Table 2.1. The sample is balanced by gender. When comparing descriptive statistics about parental background with data about the Slovak majority population that lives in close proximity to Roma, we find that the parents of subjects in our sample received on average somewhat more schooling and

[8] Originally, we planned to study also whether the ethnic minority discriminates against members of the majority group. Due to difficult access to schools with a high proportion of Roma students, which prevented us from having a large enough sample that would allow a meaningful analysis, we leave this question open for future research.

[9] Overall, in Eastern Slovakia, around 6% of primary school students leave for a selective track (after grade 5). The proportion is likely to be lower for our sample of schools, since most of this transition happens in larger towns.

[10] In the classes we study there was around 4% of Roma students. We exclude all Roma students (14) and all students who observed their choices (10) from the analyses for interpretation reasons. The results remain very similar if we keep these observations in the sample (available upon request).

[11] We obtained permission to run the experiments from the Director of Institute of Economic Studies at the Faculty of Social Sciences, Charles University in Prague, from the Dean of the Technical University of Košice and from the headmasters of participating schools. The research was also officially supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic.

have better socioeconomic position (Column 12).

2.3.2 Experimental tasks

In order to identify hostile behavior, we administered a money-burning game, which we refer to as the Joy of Destruction game (Abbink and Herrmann 2011; Abbink and Sadrieh 2009), also denoted as the JoD game. Two players received €2 each and simultaneously chose whether to pay €0.2 to reduce the counterpart's income by €1 or to keep the payoffs as they were. Since choices were made simultaneously without knowing what the counterpart did and the subjects had to pay to destroy the other's income, selfishness or sequential fairness motives cannot justify destruction. The destructive choice leads to outcomes far below the social optimum. In principle, it can be motivated by anti-social preferences (the pleasure of being nasty), an interpretation highlighted by previous work (Abbink and Sadrieh 2009), or by a hostile action triggered by beliefs about the destructive behavior of the counterpart. In the following text we will denote the choice to reduce the other's payoff in the JoD game as hostile or destructive.

We also administered a Prisoner's Dilemma game, also denoted as the PD game, in order to elicit willingness to cooperate. In this game, both players received €1.60 and simultaneously chose whether to take €0.80 from the counterpart to obtain €0.40 themselves, or to keep the payoffs as they were. While taking €0.80 is a dominant strategy for a purely selfish player, the socially optimal outcome is obtained when neither player chooses to take the other player's money, i.e. when they make cooperative choices. We denote the decision not to take the other's income in the PD game as a cooperative choice.

The subjects participated in both games, the JoD game and the PD game, in an order that was randomized across schools. In each game, participants were asked to first make an unconditional decision, i.e. choose what they wanted to do without knowing what the other player did. Subsequently, the participants were also asked to state their beliefs about the decision of the counterpart and, using the strategy method, to make two conditional decisions — for the situation when the counterpart decided to keep the payoffs unchanged, and for the situation when the counterpart decided to lower the decision maker's payoff.

2.3.3 Experimental manipulations

We have orthogonally manipulated two dimensions: the identity of the counterpart in the experimental tasks and the social environment in which the players make their decisions. First, in order to assess the extent of discrimination against Roma, we implemented SAME and OTHER conditions. In the SAME condition, the anonymous counterpart came from the majority group (i.e. was a co-ethnic), whereas in the OTHER condition, the anonymous counterpart was Roma. In both conditions, the counterpart was unknown to the decision-maker and came from a different unspecified distant school in Eastern Slovakia. The two conditions were implemented "within subject", in a random order. The counterpart's ethnicity was never mentioned during the experiment, but instead we revealed the list of 20 names of potential counterparts (ten male and ten female names). In the SAME condition the list contained typical majority names and in the OTHER condition it contained typical Roma names. The name lists were read by the experimenter prior to choosing and were included on the top of all answer sheets.

In order to match subjects with counterparts who had corresponding names, we first identified a small sample of students with typical Roma names in a different location within the same region, and we let them participate in the same set of tasks.[12] Their names were then used in the main experiments, in the list of potential counterparts in the OTHER condition, and their choices used to determine the payoff to all the (majority) subjects in the OTHER condition. Using a similar procedure, students with typical majority names took part in the tasks and their names and choices were used as counterparts in the SAME condition.

Second, we created four conditions that differed in terms of social environment of the individual choice. These treatment conditions were implemented "between subjects" — each subject took part only in one of these treatments. In the INDIVIDUAL treatment, subjects were deciding in isolation, without being observed by classmates and with no information regarding their classmates' choices. In three other treatments we exposed subjects to peers — subjects were randomly matched with two other peers from their class and were sequentially making the same decision in a randomly determined order. We let each individual decision be immediately observed by the other two peers. No communication was allowed. Subjects in the NO PEERS OBSERVED (NPO) treatment

[12]Since in the main experiment we match peers who play with an individual counterpart, in this pilot sample we implemented an extra decision-making treatment mirroring this situation — an individual was matched with three subjects.

made their decision first, without knowing what their peers would do. Subjects in OBSERVING ONE PEER (1PO) treatment made their choice second, and thus observed the choice of the preceding player in the same task before making their own decision. Subjects in OBSERVING TWO PEERS (2PO) treatment made their choice as the last player, after observing the choices of both preceding players. In all three treatments, the subjects knew their own choices would be observed by their peers. Also, to naturally motivate subjects to pay attention to the choices of peers without priming them, we implemented payoff commonality[13] — one of the three individual decisions was randomly chosen to be payoff relevant for all three matched peers.[14]

The experiment was designed such that the social environment in "field labs" reflects as closely as possible the out-of-lab social environment, in which individuals are part of a social network and naturally observe the behavior of friends and others. The participants were matched with actual peers from their class, with whom they regularly interact. The experimenter was present in the room to provide instructions and prevent communication, but far enough not to observe choices. When making choices, three matched peers sat behind each other. Although they could not communicate verbally, after each choice subjects handed over their answer sheet to the other two matched peers to read: first, the subject in NPO treatment made a choice and let 1PO and 2PO subjects read it, then 1PO made a choice and NPO and 2PO observed and lastly, 2PO made a choice and NPO and 1PO observed. This procedure was first completed for unconditional choices, and then repeated for beliefs and conditional choices. The full experimental protocol is included in the Supplementary materials.

In all three manipulations when peers are present, the decision environment and incentives are the same, the only difference is the number (0,1,2) and type (hostile vs. non-hostile in the JoD game, cooperative vs. non-cooperative in the PD game) of signals a subject is exposed to prior to his choice. Given that matching and order was randomly determined, such a design allows us to identify the influence of peers by comparing choices

[13]These design features are inspired by Charness, Rigotti, and Rustichini (2007) who study how salience of group membership artificially created in a laboratory affects in-group preferential treatment. While we use similar techniques (being observed by others and a common payoff), our design differs in two ways. First, we measure differential treatment based on an existing group attribute (ethnicity). Second, we focus on how the actions of peers, with whom a decision-maker frequently interacts in real life, affect individual behavior.

[14]Another important issue which we study in a separate cross-country study, is how group decision-making affects the prevalence of hostility. Besides the manipulations described above we have also implemented an additional treatment in which subjects made choices jointly as a group. In contrast, in this paper we study the influence of peers on individual behavior.

of subjects exposed to a different number and type of signals with the choices in the NO PEERS OBSERVED (NPO) treatment. If there are no peer effects in hostility, then the decisions of 1PO and 2PO players should be unrelated to the choices of the preceding player/s and thus similar to choices in NPO treatment. Observing that players with hostile predecessors are more likely to be hostile themselves would indicate that hostile behavior has a tendency to spread. Ultimately, we aim to explore interaction effects, i.e. whether peers are more influential in the OTHER as compared to the SAME condition.

This design has two limitations. First, due to payoff commonality the comparison of INDIVIDUAL and NPO may capture the effects of social preferences towards their own peers, in addition to the effect of being observed, and we discuss this possibility when interpreting our findings. Second, it would be interesting to know how observing a peer's unconditional decision whether to harm affects not only unconditional decisions, but also beliefs about the actions of the counterpart and conditional choices. This would be possible in a different setup, in which unconditional choices are made public to the other two peers, while at the same time beliefs and conditional choices are reported in private such that they do not affect the unconditional choices of the subjects deciding later. Implementing such a procedure in a school setting would be logistically difficult and would seem less natural, since each subject would have to leave the other two matched peers after making the unconditional decision and return after reporting beliefs and conditional choices in private. Additionally, all four tasks (unconditional decision, beliefs, and two conditional decisions) would need to be explained to the subjects at once, prior to making choices, making the decision situation cognitively more demanding. Given this methodological tradeoff, we opted for a simpler design, in which subjects stay in one place during the whole experiment, make all decisions in sequence while being observed by two matched peers, and in each moment focus on a single decision, although such a design does not allow for estimating the effects on beliefs and conditional decisions.[15]

After the experimental tasks we collected data about a set of observable characteristics of participants and their family background. All these characteristics vary little across the experimental treatments, indicating that randomization was successful (Columns 2–6 of Table 2.1). Similarly, exposure to the hostile behavior of peers is also unrelated to observable characteristics (Columns 8–10).

[15]This is because prior to eliciting beliefs and conditional decisions, subjects received many signals, some of them likely affected by their own choices (and thus endogenous). Still, we let subjects make these decisions, in order to make the set of choices and probability of making payoff-relevant choices the same as in the INDIVIDUAL treatment.

2.3.4 Procedures

After a general introduction, the students randomly picked an ID number, which determined in which of the four treatments they were allocated (INDIVIDUAL, NPO, 1PO and 2PO) and, in the last three treatments, with whom they were matched. The treatment and matched peers remain the same for both tasks and across SAME and OTHER conditions. The experiment was implemented in two separate rooms, one for INDIVIDUAL treatment and the second one for NPO, 1PO and 2PO. Experimental instructions were provided by five experimenters from the majority ethnic group, who were randomly allocated to rooms. The results are robust to controlling for experimenter fixed effects. To avoid communication about experimental tasks prior to participating, all subjects from each class participated in the experiment at the same time and all sessions within each school were implemented in a single day. Each session lasted around 1.5 hours.

To ensure understanding, the tasks were explained in detail, using visual aids to illustrate options and payoffs. Before making choices, participants were asked four control questions about the payoff consequences of their actions as well as those of their counterpart. Comprehension was generally high. In the JoD game and PD game, subjects answered all four questions correctly at the first attempt in 82% and 78% of cases, respectively. The results are robust to excluding observations with imperfect understanding.

While we deliberately manipulated the degree of anonymity in choices with respect to peers, we took several steps to ensure that choices were anonymous with respect to the experimenters, teachers and parents. First, subjects submitted all answers under their experimental IDs and were never asked for their names. Second, the experimenter who explained the tasks to the subjects could not observe the decisions made, as all answer sheets and questionnaires were submitted privately into a box located in the corner of the classroom. The answer sheets were processed and payments were later administered by a different person. Third, subjects were assured that the experimenters would not share information about decisions and resulting earnings with the teachers and parents of the participants. As a result, the choices in the INDIVIDUAL treatment were not observed, directly or indirectly, by anybody and in the NPO, 1PO and 2PO subjects were observed exclusively by the two matched peers (who could, potentially, spread the information further).

In each identity condition (SAME or OTHER) the experimenter explained that only one choice (out of six, three in the JoD game and three in the PD game) would be

randomly selected to be payoff relevant, by drawing a ball out of a bag at the end of the experiment. Thus, subjects were paid for two choices in total. This limits the scope for hedging across choices and avoids the problem that could arise if the subjects made choices with an eye on total payoff received by an experimental counterpart, instead of payoffs in the given task. Similarly, it was randomly determined whether the choice of subject in NPO, 1PO or 2PO was payoff relevant for all three matched players.

Experimental payoffs were denoted in real money. Subjects received rewards in the form of credit to order items from an experimental store, which contained 48 items, ranging from sweets, snacks and drinks to stationery, stickers and bracelets, to satisfy a variety of tastes. All items were priced using retail prices. Prior to the experiments participants were provided with a "store catalog", in order to learn about items (depicted with pictures) and prices. After the experiments they selected their preferred items, which were later distributed to schools in sealed bags marked with experimental IDs.

2.4 Results

In this section we start by describing choices in the NPO condition, i.e. in the situation where peers observe the decision-maker's choices, but the decision-maker does not observe the choices of peers. As a next step, we study choices in 1PO and 2PO conditions, in which subjects are exposed to the choices of peers (in addition to being observed by them), in order to understand the susceptibility to follow peer behavior.

We find that a non-negligible proportion of subjects in NPO chooses to destroy in the JoD game (45%). There is no evidence of discrimination — if anything, subjects are less destructive to the ethnic minority: 47% chose to destruct in SAME, while 42% in OTHER, and the difference is not significant statistically (Column 1 in Panel A of Table 2.2). A similar result also arises in the PD game: the proportion of cooperative choices is almost identical across conditions, 28% in SAME and 30% in OTHER (Column 1 in Panel B).

Observation 1: We do not find evidence of ethnic discrimination when individuals make choices in front of their peers (but without observing their behavior). This holds both when subjects decide whether to cause harm and whether to cooperate.

Next, we explore the influence of peer behavior in the JoD game. We find that individuals follow the destructive behavior of classmates, especially in the OTHER condition,

and we document this pattern for choices of subjects in both 1PO and 2PO conditions. In 1PO (Columns 5 and 6 in Panel A of Table 2.2), when subjects prior to their own choice observe one peer who chooses to be non-destructive, the frequency of destructive choices in OTHER is 19%. When the preceding player is destructive, the frequency sharply increases to 77%. Peers also influence behavior in SAME, but to a lesser extent: the prevalence of destructive behavior increases from 23% when a peer is non-destructive to 51% when he is destructive. Put differently, in a regression analysis we find a strong positive interaction effect between making their own choice after the destructive behavior of classmates and being in the OTHER condition on the likelihood of choosing a destructive action (Column 1 of Panel B in Table 2.3). As a result, discrimination against the ethnic minority arises when participants observe classmates' hostility and the gap is large in magnitude (29 percentage points) and highly statistically significant (Column 4 in Panel A of Table 2.5), while we find no such discrimination when participants observe their peer being non-hostile (Column 5).

We find very similar interaction effects for subjects who made choices after observing the choices of two peers (2PO). We first compare the behavior of (a) individuals who were exposed to observing the destructive behavior of two classmates and (b) individuals who were not (those who observed consistently non-destructive peers or mixed behavior of preceding players). The difference is 70 percentage points (88% and 18%, resp.) in OTHER and 38 percentage points (67% and 29%, resp.) in SAME (Columns 8 and 9 in Panel A of Table 2.2). The difference in effects of destructive peer behavior is statistically significant across conditions (Column 1 in Panel C of Table 2.3). Two further results are noteworthy. In line with intuition, the effects of observing two hostile peers (2PO) are somewhat larger than the effect of observing one hostile peer (1PO) in both conditions, although the differences are not significant statistically. Interestingly, receiving a mixed signal (one peer destructive, one peer non-destructive) is not enough to generate ethnic discrimination — the prevalence of hostile choices is somewhat smaller in OTHER than in SAME (Panel A of Table 2.7).

Given the similarity of effects in 1PO and 2PO, in Panel A of Table 2.3 we pool choices in these two treatments and estimate the effects of being exposed to observing consistently destructive behavior of peers (one destructive peer in 1PO and two destructive peers in 2PO), in order to increase the power of our estimates. The interaction effect of destructive behavior of peers and OTHER (Column 1 in Panel A) is again large in magnitude and highly significant statistically (35 percentage points, p-value=0.001), and

substantial discrimination between SAME and OTHER conditions occurs when peers are destructive (27 percentage points, p=0.002, Column 2 in Panel A of Table 2.5).

Observation 2: Peers are very influential in shaping individual willingness to engage in destructive behavior, especially when the counterpart is the ethnic minority.

Next, we explore whether the greater influence of peers in OTHER is due to a greater susceptibility to following destructive behavior or due to susceptibility to following non-destructive behavior of peers. Behavior in NPO offers a natural comparison group, since the decision-making environment is similar to 1PO and 2PO — the only difference being the signal the decision-maker received prior to choosing — and in the NPO condition we find little difference across OTHER and SAME conditions. Figure 2.2 and Figure 2.3 reveal a clear picture. Relative to the choices of the subjects in NPO, subjects in 1PO and 2PO who were exposed to non-destructive peers are around 20–25 percentage points less likely to be destructive themselves. The magnitude of this effect is very similar for SAME and OTHER conditions (Columns 5 and 6 in Panel A of Table 2.3). Strikingly, however, the susceptibility to following the destructive behavior of peers is specific for the OTHER condition. Specifically, in OTHER subjects who observed peer/s to be destructive are 41 percentage points more likely to be destructive themselves (Column 6) compared to subjects in NPO. In contrast, we find virtually no effect in SAME condition (Column 5). Again, the interaction effect is large in magnitude (32 percentage points) and highly significant statistically. We find similar interaction effects (33 p.p. and 27 p.p.) and arrive at the same conclusions when analyzing the choices separately in 1PO and 2PO (Panel B and Panel C, respectively). Together, these results suggest that a greater influence of peers on individual destructive behavior towards the ethnic minority originates in a greater readiness to follow peers when they are destructive, but not by differences in readiness to follow non-destructive behavior.

Observation 3: The greater influence of peers on destructive behavior towards an ethnic minority, as compared to co-ethnics, is driven by a greater susceptibility to following destructive behavior and not by differences in susceptibility to following non-destructive behavior. As a result, ethnic discrimination arises when peers choose to harm, while we find virtually no discrimination when they are peaceful.

The results from the PD game provide further support for the interpretation that the magnified readiness to follow observed behavior in OTHER as compared to SAME condition is specific for unambiguously destructive behavior, but does not necessarily

apply to other types of behaviors. Using the same specifications as before, we find that peers again matter in intuitive ways (Columns 4–6 in Panel A of Table 2.4): observing classmate/s cooperating increases individual willingness to cooperate (by around 30 percentage points) and observing classmates defecting and acting selfishly reduces the likelihood of cooperation (by around 13 percentage points). Importantly, however, the propensity to follow others is very similar in SAME and OTHER conditions. Thus, we find virtually no discrimination in cooperative behavior across SAME and OTHER conditions, independently to whether peers choose to cooperate or not (Panel B of Table 2.5).

Observation 4: Peers are influential in shaping cooperative behavior. The magnitude of peer effects is similar independently to the identity of the counterpart.

As a next step, we analyze the results from the INDIVIDUAL treatment (Column 11 in Table 2.2). We find that the prevalence of destructive choices in the JoD game is 30% in both conditions. Also, the prevalence of cooperative choices in the PD game is 31% in SAME and 32% in OTHER. Thus, when individuals make choices in isolation from others we find no evidence of discrimination, both in the destructive as well as cooperative type of interaction. This is interesting because this decision environment has been widely used in previous studies to assess the prevalence of discrimination. However, as we have demonstrated above, social context matters, and relying on decisions in this environment only may lead to underestimating the risk of inter-ethnic tensions.

It is also noteworthy that the proportion of destructive choices in the JoD game is higher in NPO compared to INDIVIDUAL independently to the identity of the counterpart (Column 12 in Panel A of Table 2.2), in line with previous work in social psychology on "discontinuity effect" (Wildschut et al. 2003), which suggests that people in groups behave in a more hostile and competitive way when compared to individuals. This effect may be due to the effect of being observed by peers, or it can in principle be due to spiteful preferences towards own peers, since the individual decision impacts not only their own payoff but also the payoff of the other two peers (payoff commonality). We do not find any differences in the PD game (Panel B).

The INDIVIDUAL treatment provides clear measures of individual prior beliefs and preferences (conditional choices). In principle, since subjects were randomly allocated to treatments, these measures should be informative about prior beliefs and preferences of subjects in NPO, 1PO and 2PO treatments with which they entered social interactions

with peers. We find significant differences in beliefs about the behavior of counterparts in the JoD game across identity of the counterpart (Panel A of Table 2.6). Specifically, in the SAME condition the proportion of subjects who expect their counterpart to be destructive is 35%, while it is 49% in the OTHER condition. This indicates both greater expected hostility of the ethnic minority, in line with the idea that the ethnic minority is seen as a greater threat, as well as greater uncertainty, since beliefs are very close to 50% in OTHER.

In the PD game we find that beliefs about behavior of the counterpart are the same in both conditions — 34% of subjects expect the counterpart to cooperate (Panel B).

Observation 5: In the Joy of Destruction game subjects expect the ethnic minority to be more destructive than the majority group, while there is no such difference in beliefs about cooperative behavior in the Prisoner's Dilemma game.

Table 2.6 reports results for conditional choices in both tasks. In the JoD game we find somewhat more destructive behavior in OTHER compared to SAME, although observed differences are not significant statistically. Conditional on the counterpart choosing the destructive strategy, the likelihood of choosing the destructive strategy as well is 51% in SAME and 58% in OTHER (p-value=0.33), while when the counterpart chooses not to destruct, the likelihoods are 28% and 33%, respectively (p-value=0.37). In the PD game, conditional on defection of the counterpart, the likelihood of cooperation is 16% in SAME and 9% in OTHER, and this difference is marginally significant statistically (p-value=0.09). When the counterpart cooperates, the likelihoods of cooperation are similar across conditions (28% and 30%, resp.). In sum, we find rather moderate evidence of preference-based discrimination when subjects are isolated from their peers.

Robustness checks

We now report a series of robustness analyses of the main finding: the greater susceptibility to following peers when deciding whether to engage in destructive behavior towards an ethnic minority as compared to co-ethnic. First, sub-group analyses do not suggest that the main effect would be driven by a particular demographic group. Tables 2.8 and 2.9 report the main results separately for male and female decision-makers. We find that, overall, the patterns are qualitatively similar for both groups, but the magnitude and statistical significance is larger for male decision-makers. Also, the main effect is not driven by subjects with low socioeconomic background — we observe qualitatively simi-

lar results for the sub-sample of subjects who have at least one parent with a university degree (Tables 2.10 and 2.11). Second, we test for the potential role of the inability to understand the task. We exclude 18% of observations in which subjects did not answer correctly at the first attempt at least one comprehension question (out of four) we administered before each task and repeat the main analysis: the results are robust (Tables 2.12 and 2.13). Third, we control for experimenter fixed effects, order of the task, and order of the OTHER vs. SAME condition, and find the results to be robust (Tables 2.14 and 2.15). Further, since we have implemented OTHER and SAME conditions within subjects, we test whether results are robust if we focus on choices in the condition that was implemented first (OTHER or SAME), thus mimicking a "between-subject" design. The results are robust and statistically significant (Tables 2.16 and 2.17). Last, while in the main tables we use OLS, the results are robust to using probit (available upon request).[16]

2.5 Interpretation

In this section we consider which theoretical mechanisms can explain the main findings: (i) greater susceptibility to following destructive actions of peers in the OTHER condition compared to the SAME condition and (ii) no differences in following across conditions when peers act non-destructively in the JoD game and cooperatively in the PD game.

The first type of explanation is based on parochial response to conflict, in the spirit of the evolutionary models (Choi and Bowles 2007; Bowles 2008). The pattern is consistent with the idea that hostile behavior of peers signals a threat of ethnic conflict, which in turn triggers parochial motivations. Since parochial motivations are predicted to cement in-group cohesion, either by greater within-group altruism or expectation of future punishment, the actions of peers should become more contagious. Parochialism may also trigger out-group hate and in such case the hostile actions of peers would affect individual willingness to do harm not only because of elevated motivations to conform with in-group members, but also due to the intensified dislike of the ethnic minority. Also in line with this explanation, we do not find any differences in following across conditions when the decision-maker observes choices of peers in the PD game or non-destructive choices in the JoD game, i.e. choices which do not signal a threat of ethnic conflict and

[16]Despite having a binary dependent variable, we used OLS in our main analysis because of concerns about interactions effects in probit regressions.

thus parochialism is not predicted to intensify. The explanation is also supported by prior beliefs, which suggest that the ethnic minority is seen as a greater threat relative to the majority (co-ethnics).[17]

Next, we consider information-based explanations.[18] First, social learning from observing others and associated herding is predicted to be the larger, the weaker are individual priors about behavior of experimental counterpart.[19] The measure of prior beliefs elicited in the INDIVIDUAL condition suggests that the participants are more uncertain about actions of Roma than about actions of co-ethnics in the JoD game. Specifically, 35% expect the counterpart to be destructive in the SAME condition, while it is 49% in the OTHER condition. Such differences in beliefs are not found in the PD game. Social learning thus predicts signals from peers to be more influential in the OTHER relative to the SAME condition in the JoD game, but not in the PD game. However, it struggles to explain why only hostile behavior of peers in the JoD game (and not non-destructive actions) turned out to be more influential in the OTHER as compared to the SAME condition. To explain all our findings, this information-based mechanism would require a positive skew of the distribution of individual beliefs about the behavior of Roma in the JoD game (i.e. harboring suspicion about the hostility of Roma).

Second, the information-based mechanism could be due to uncertainty about social norms guiding behavior towards others. If participants were more uncertain about the normatively right way of behaving towards Roma than towards co-ethnics, the observed choices of peers would be predicted to be more influential in the OTHER than in the SAME condition. Note that also in this case it is hard to explain why we observe the asymmetry in following only when peers cause harm, and not when they act peacefully.

In sum, the full set of findings is consistent with harmful actions of peers triggering parochial psychological response. At the same time, we cannot rule out that our results

[17]Interestingly, recent research in social psychology documents an intensified readiness to associate threat with individuals of other ethnicity compared to individuals of one's own ethnicity (Olsson et al. 2005; Golkar, Björnstjerna, and Olsson 2015; Mallan, Sax, and Lipp 2009; Navarrete et al. 2009). In these experiments, the researchers measure fear learning by physiological reaction to aversive stimulus, which participants receive when observing images of faces of people of own vs. other ethnicity.

[18]For information-based models of contagion, based on observing the past decisions of others, see Banerjee (1992) and Bikhchandani, Hirshleifer, and Welch (1992). Bikhchandani, Hirshleifer, and Welch (1998) provide an overview of this literature.

[19]Beliefs about the actions of the counterpart are predicted to matter in both the PD game, since conditional cooperation is common behavior in this type of experiment, as well as in the JoD game, since hatred and associated destructive behavior are often considered a defensive emotional response to feelings of being threatened by somebody (Baumeister 1995; Dozier 2002; Glaeser 2005) and thus beliefs about the more destructive nature of members of a certain group could motivate pre-emptive aggressive action.

are driven by some form of uncertainty, specific for destructive interactions with Roma.

2.6 Conclusions

In order to understand the sources of inter-group conflict, the exploration of behavior towards dissimilar ethnic groups has been central across disciplines in social sciences for decades. At the same time, it is well established that the actions of other people within their own social group can greatly affect individual behavior. Yet, there is no direct evidence as to what extent such social influences matter when individuals make choices whether to engage in harming a dissimilar ethnic group, and thus how contagious ethnic hostility is. This is what we provide.

Our findings are subtle and telling at the same time. On one hand, we find that subjects do not discriminate against the ethnic minority when choices are performed in isolation from peers or when individuals are exposed to observing the peaceful behavior of peers, which is an encouraging finding in light of the widespread concern about the pervasive nature of ethnic discrimination. On the other hand, however, our results demonstrate that an individual decision whether to be destructive or not towards the ethnic minority is very fragile and social influences matter a great deal. We find that individual tendency to follow the destructive behavior of peers is amplified when the subject of hostility is a member of the ethnic minority, as compared to a co-ethnic. As a consequence, hostile actions towards the ethnic minority tend to spread and ethnic bias in hostility arises. The findings are consistent with elevated parochialism — implying greater in-group cohesion or out-group hostility — in response to signals of threat of ethnic conflict. We also discuss potential explanations based on social learning.

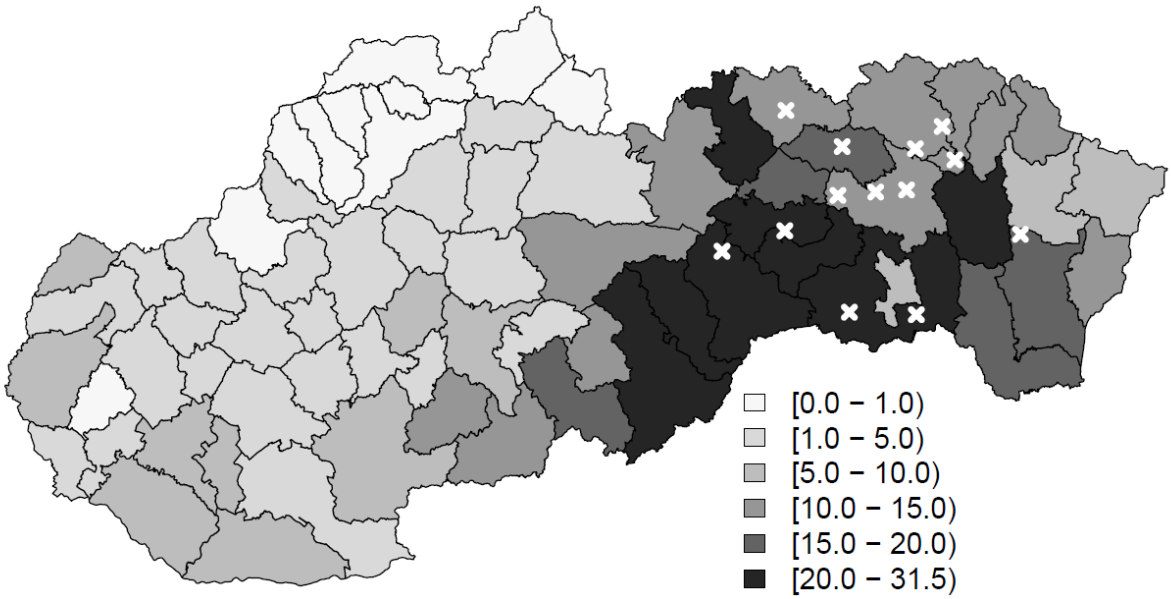
Establishing strong contagion when individuals see others doing harm to members of another ethnic group may illuminate why ethnic hostilities of masses can spread quickly, even in those societies, in which there are few visible signs of systematic inter-ethnic hatred (Fearon and Laitin 2000; Basu 2005) and why "entrepreneurs of hatred" (i.e. individuals who could benefit from causing social unrest and conflict), perhaps aware of inflammability of ethnic hostility, often choose other ethnic groups as targets of their aggressive political campaigns. Glaeser (2005) provides a political economy model, which assumes the easy spreading of hatred towards minority groups and explores when political entrepreneurs are motivated to supply the masses with hate-creating stories.

The results are potentially relevant for policy, too. Since ethnically-motivated hos-

tility has negative externalities which go beyond the immediate victim, by potentially threatening social cohesion in a community the results can help to explain why many societies have found it desirable to institute hate crime laws (a special category of legislation aimed to punish, particularly severely, those offences which are motivated by ethnicity, religion or other group attributes of the victim). In addition, early diagnoses seem to be of vital importance, since our results suggest that although the readiness to be hostile may be quite latent and thus harmless in peaceful times, it may gain importance when ethnic tensions arise.

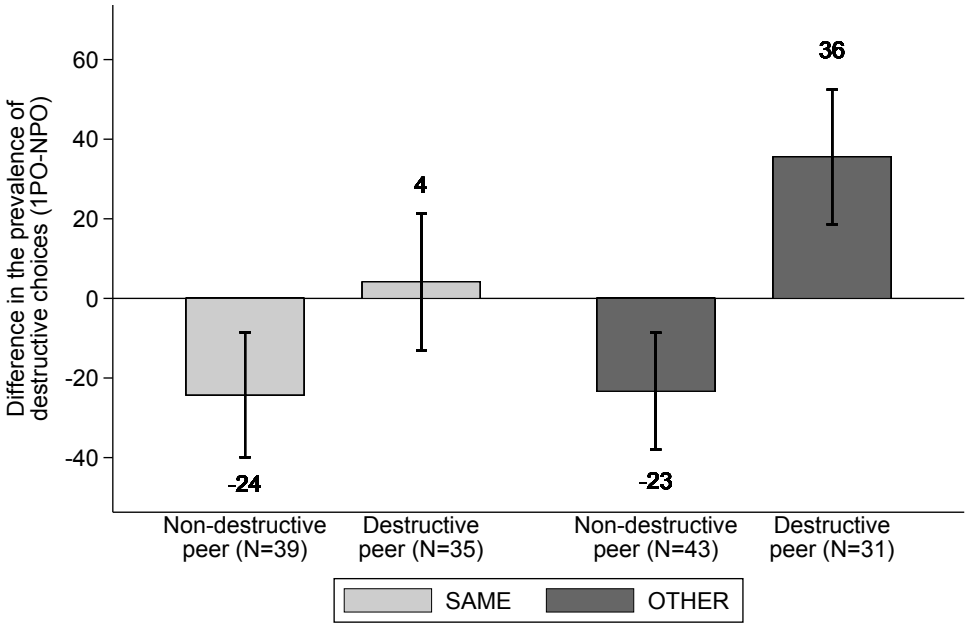
This is the first experiment on social contagion of ethnic hostility and naturally it raises as many questions as it answers. Our experimental setup was designed to explore the social contagion of doing harm to an ethnic minority among individuals who know each other and can observe each other's choices, mimicking many real-life situations. It is a fruitful area for future research to explore whether our findings generalize to situations where individuals receive signals about the behavior of peers anonymously, and thus when expectations of future punishment cannot drive a decision to conform. Another direction is to study the role of social distance and previous contact, for instance, whether the harmful actions of individuals outside of the immediate social network trigger a similar behavioral response as those of peers. Furthermore, in light of the recent migration wave from the Middle East, Afghanistan and Africa to Europe and anecdotal evidence suggesting the quick spreading of fear among ordinary people in many recipient countries, it would be interesting to investigate whether the degree of contagion of hostile behavior is related to previous contact with a given ethnic group. We believe our experimental design, which combines exogenous variation in ethnic identity and social context, adds to the portfolio of empirical tools that can help to make progress towards better understanding of these important issues.

Figure 2.1: Location of the study (map of Slovakia)



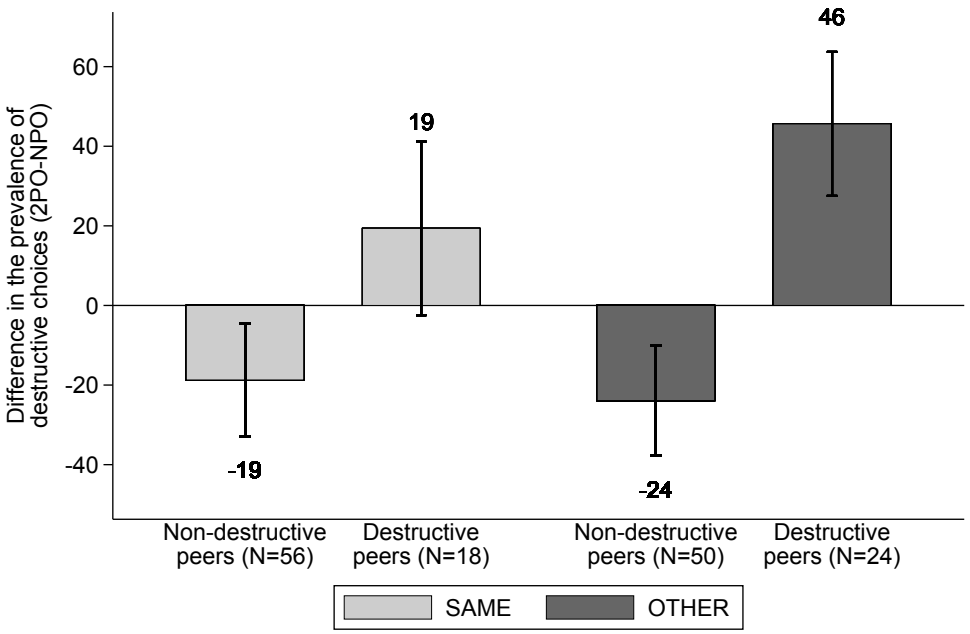
Notes: Percentage share of Roma in total district population, using data from Mušinka et al. (2014). White crosses indicate the location of schools where the experimental data were collected.

Figure 2.2: Effect of observing the action of one peer, Joy of Destruction game



Notes: Effect of observing the action of one peer on individual destructive behavior in the Joy of Destruction game. The figure reports the difference (in percentage points) between the prevalence of destructive choices in the OBSERVING ONE PEER (1PO) and NO PEERS OBSERVED (NPO) condition, by the behavior of the preceding player in the 1PO condition. "Non-destructive peer" indicates that the subject in the 1PO condition observed non-destructive behavior of a preceding player, while "Destructive peer" indicates that the subject observed destructive behavior of a preceding player. SAME and OTHER indicate whether the subjects are deciding in the SAME condition (Majority partner) or in the OTHER condition (Roma partner). Bars indicate 90% confidence intervals.

Figure 2.3: Effect of observing the actions of two peers, Joy of Destruction game



Notes: Effect of observing the actions of two peers on individual destructive behavior in the Joy of Destruction game. The figure reports the difference (in percentage points) between the prevalence of destructive choices in the OBSERVING TWO PEERS (2PO) and NO PEERS OBSERVED (NPO) condition, by the behavior of the preceding players in the 2PO condition. "Destructive peers" indicates that the subject in the 2PO condition observed that both preceding players decided to destruct. "Non-destructive peers" indicates that the subject in the 2PO condition observed that one or both preceding players decided not to destruct. SAME and OTHER indicate whether the subjects are deciding in the SAME condition (Majority partner) or in the OTHER condition (Roma partner). Bars indicate 90% confidence intervals.

Table 2.1: Descriptive statistics and randomization checks

	Means across treatments							Means by behavior of peers in the Joy of Destruction game				Slovakia
	Whole sample	IND.	NPO	1PO	2PO	F-test (p-val.)	N	1PO & 2PO observing destructive peers	1PO & 2PO observing non-destructive peers	t-test (p-val.)	N	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(9)	(8)	(10)	(11)	(12)#
Female	0.50	0.51	0.43	0.51	0.53	0.65	326	0.46	0.55	0.16	294	
Age	13.79	13.71	13.95	13.78	13.74	0.25	323	13.76	13.76	0.95	292	
Cognitive skills (0–4)	2.96	2.96	3.03	2.89	2.95	0.89	326	2.87	2.95	0.56	294	
Number of siblings	1.81	1.80	1.76	1.71	1.97	0.74	278	1.90	1.80	0.59	246	
Mother unemployed	0.22	0.20	0.18	0.22	0.28	0.57	294	0.24	0.25	0.88	256	0.40
Father unemployed	0.10	0.13	0.08	0.07	0.12	0.44	284	0.07	0.11	0.28	240	0.27
Mother with high school	0.77	0.79	0.74	0.73	0.85	0.37	261	0.83	0.75	0.20	236	0.74
Mother with university	0.21	0.20	0.24	0.27	0.13	0.25	261	0.17	0.23	0.29	236	0.07
Father with high school	0.80	0.76	0.84	0.82	0.82	0.68	249	0.81	0.83	0.76	232	0.76
Father with university	0.18	0.19	0.15	0.18	0.18	0.90	249	0.19	0.17	0.76	232	0.07
Family owns a car	0.85	0.88	0.85	0.88	0.77	0.29	315	0.83	0.82	0.89	282	0.66
Family owns a computer	0.97	0.96	0.96	1.00	0.94	0.30	320	0.98	0.97	0.51	286	0.63
Family owns a TV	0.98	0.99	0.97	1.00	0.96	0.23	319	0.98	0.98	0.85	284	0.99
Family owns a tablet	0.32	0.24	0.37	0.37	0.32	0.31	260	0.35	0.34	0.89	238	

Notes: Descriptive statistics of the sample. Columns 2–5 present means across the INDIVIDUAL, NO PEERS OBSERVED (NPO), OBSERVING ONE PEER (1PO) and OBSERVING TWO PEERS (2PO) treatments. Experimental balance is tested in Column 6 using an F-test. Columns 8 and 9 compare subjects in the 1PO and 2PO conditions who observed the destructive behavior of preceding players in the Joy of Destruction game with those who did not; results of a t-test are presented in Column 10. Dummy variables "Mother Unemployed", "Father Unemployed" are equal to one if the subject reported that the parent does not have a job and equal to zero if the subject reported that the parent has a job. Education of parents is summarized using dummy variables identifying whether the highest completed education level is university, high school, or "lower" education level (omitted group). Variable "Cognitive skills" is equal to the number of correctly answered Raven's Progressive Matrices (4=max). For all variables, the values are missing for unspecified answers and for "I do not know" answers; see N in Columns 7 and 11 for the number of non-missing values. The data for Slovakia presented in Column 12 come from the UNDP & FRA Survey (2011) which is based on a sample of adults living in the same areas as Roma.

Table 2.2: Means, across experimental manipulations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Destructive choice in the Joy of Destruction game												
Sample	NPO		1PO and 2PO			1PO			2PO		INDIVIDUAL	
	All	Destructive peer(s)	Non-destructive peer(s)	P.p. diff. (2)-(3); (p-value)	Destructive peer	Non-destructive peer	P.p. diff. (5)-(6); (p-value)	Destructive peers	Non-destructive peers	P.p. diff. (8)-(9); (p-value)	All	P.p. diff. (1)-(11); (p-value)
OTHER	0.42	0.82	0.18	64 (0.00)	0.77	0.19	43 (0.00)	0.88	0.18	70 (0.00)	0.30	12 (0.06)
SAME	0.47	0.57	0.26	30 (0.00)	0.51	0.23	28 (0.01)	0.67	0.29	38 (0.00)	0.30	17 (0.01)
P.p. diff. OTHER-SAME; (p-value)	5 (0.51)	25 (0.00)	-8 (0.19)		26 (0.03)	-4 (0.62)		21 (0.10)	-11 (0.20)		1 (0.88)	
N	148	108	188		66	82		42	106		210	
Panel B: Cooperative choice in the Prisoner's Dilemma game												
Sample	NPO		1PO and 2PO			1PO			2PO		INDIVIDUAL	
	All	Cooperative peer(s)	Non-cooperative peer(s)	P.p. diff. (2)-(3); (p-value)	Cooperative peer	Non-cooperative peer	P.p. diff. (5)-(6); (p-value)	Cooperative peers	Non-cooperative peers	P.p. diff. (8)-(9); (p-value)	All	P.p. diff. (1)-(11); (p-value)
OTHER	0.30	0.56	0.20	36 (0.00)	0.45	0.19	26 (0.02)	0.8	0.20	60 (0.00)	0.32	-3 (0.71)
SAME	0.28	0.59	0.16	44 (0.00)	0.52	0.15	37 (0.00)	0.73	0.16	57 (0.00)	0.31	-3 (0.66)
P.p. diff. OTHER-SAME; (p-value)	1 (0.86)	-3 (0.80)	4 (0.39)		-7 (0.65)	4 (0.57)		7 (0.70)	4 (0.52)		1 (0.88)	
N	148	64	232		43	105		21	127		210	

Notes: Means. Panel A reports the prevalence of destructive behavior in the Joy of Destruction game. Panel B reports the prevalence of cooperative behavior in the Prisoner's Dilemma game. OTHER and SAME indicate whether the subjects are deciding in the OTHER condition (Roma partner) or in the SAME condition (Majority partner). Column 1 reports behavior in the NO PEERS OBSERVED (NPO) condition. Columns 2-3 pool observations from the OBSERVING ONE PEER (1PO) and OBSERVING TWO PEERS (2PO) conditions. Columns 5-6 report choices in the 1PO condition and Columns 8-9 in the 2PO condition. In Panel A, "Destructive peer(s)" indicates that all preceding players matched with a given subject (one player in 1PO and two players in 2PO) made a destructive choice. "Non-destructive peer(s)" indicates that in 1PO the preceding player made a non-destructive choice and in 2PO one or both preceding players made a non-destructive choice. In Panel B, "Cooperative peer(s)" indicates that all preceding players matched with a given subject (one player in 1PO and two players in 2PO) made a cooperative choice. "Non-cooperative peer(s)" indicates that in 1PO the preceding player made a non-cooperative choice and in 2PO one or both preceding players made a non-cooperative choice. Column 11 reports behavior in the INDIVIDUAL condition. All differences are presented in percentage points and tested using a Chi-square test.

Table 2.3: The effect of peer behavior on destructive choices (Joy of Destruction game)

Dependent variable	**Destructive choice in the Joy of Destruction game**					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Effect of one or two peers						
Sample	1PO and 2PO			NPO, 1PO and 2PO		
	All	SAME	OTHER	All	SAME	OTHER
OTHER	-0.08			-0.05		
	(0.06)			(0.08)		
Destructive peer(s)	0.29***	0.30***	0.64***	0.09	0.09	0.41***
	(0.08)	(0.08)	(0.06)	(0.09)	(0.09)	(0.08)
Non-destructive peer(s)				-0.21***	-0.20***	-0.23***
				(0.07)	(0.07)	(0.07)
OTHER*Destructive peer(s)	0.35***			0.32***		
	(0.10)			(0.12)		
OTHER*Non-destructive peer(s)				-0.02		
				(0.10)		
Observations	294	147	147	442	221	221
Panel B: Effect of observing one peer						
Sample	1PO			NPO and 1PO		
	All	SAME	OTHER	All	SAME	OTHER
OTHER	-0.04			-0.05		
	(0.09)			(0.08)		
Destructive peer	0.28**	0.29**	0.61***	0.04	0.04	0.36***
	(0.11)	(0.11)	(0.09)	(0.10)	(0.10)	(0.09)
Non-destructive peer				-0.25***	-0.24***	-0.22***
				(0.09)	(0.09)	(0.08)
OTHER*Destructive peer	0.32**			0.33**		
	(0.14)			(0.14)		
OTHER*Non-destructive peer				0.02		
				(0.12)		
Observations	146	73	73	294	147	147
Panel C: Effect of observing two peers						
Sample	2PO			NPO and 2PO		
	All	SAME	OTHER	All	SAME	OTHER
OTHER	-0.11			-0.05		
	(0.08)			(0.08)		
Destructive peers	0.37***	0.38***	0.69***	0.19	0.19	0.45***
	(0.13)	(0.13)	(0.09)	(0.13)	(0.13)	(0.09)
Non-destructive peers				-0.18**	-0.17**	-0.24***
				(0.08)	(0.09)	(0.08)
OTHER*Destructive peers	0.32**			0.27*		
	(0.16)			(0.16)		
OTHER*Non-destructive peers				-0.05		
				(0.12)		
Observations	148	74	74	296	148	148

Notes: OLS, robust standard errors in parentheses. *** denotes p<0.01, ** p<0.05 and * p<0.1. Panel A reports results together for subjects in the OBSERVING ONE PEER (1PO) and OBSERVING TWO PEERS (2PO) conditions. Panel B and Panel C report results separately for the 1PO and 2PO condition, respectively. OTHER is a dummy variable equal to one if the subject made a decision in the OTHER condition (Roma partner) and zero for choices in the SAME condition (Majority partner). See Table 2 for definitions of the variables "Destructive peer(s)" and "Non-destructive peer(s)." In Columns 1-3 the subjects in the NO PEERS OBSERVED (NPO) are excluded and the omitted group is "Non-destructive peer(s)". In Columns 4-6 the omitted group is the NPO condition. In all Columns of all Panels, we control for gender and school grade.

Table 2.4: The effect of peer behavior on cooperative choices (Prisoner's Dilemma game)

Dependent variable	**Cooperative choice in the Prisoner's Dilemma game**					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Effect of one or two peers						
Sample	1PO and 2PO			NPO, 1PO and 2PO		
	All	SAME	OTHER	All	SAME	OTHER
OTHER	0.05			0.01		
	(0.05)			(0.07)		
Cooperative peer(s)	0.48***	0.48***	0.39***	0.33***	0.33***	0.28**
	(0.09)	(0.09)	(0.10)	(0.10)	(0.10)	(0.11)
Non-cooperative peer(s)				-0.14**	-0.14**	-0.11*
				(0.06)	(0.06)	(0.06)
OTHER*Cooperative peer(s)	-0.09			-0.05		
	(0.13)			(0.15)		
OTHER*Non-cooperative peer(s)				0.03		
				(0.09)		
Observations	294	147	147	442	221	221
Panel B: Effect of observing one peer						
Sample	1PO			NPO and 1PO		
	All	SAME	OTHER	All	SAME	OTHER
OTHER	0.04			0.01		
	(0.07)			(0.07)		
Cooperative peer	0.41***	0.41***	0.29**	0.26**	0.26**	0.17
	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.13)
Non-cooperative peer				-0.14**	-0.14**	-0.11
				(0.07)	(0.07)	(0.08)
OTHER*Cooperative peer	-0.12			-0.10		
	(0.18)			(0.18)		
OTHER*Non-cooperative peer				0.03		
				(0.10)		
Observations	146	73	73	294	147	147
Panel C: Effect of observing two peers						
Sample	2PO			NPO and 2PO		
	All	SAME	OTHER	All	SAME	OTHER
OTHER	0.05			0.01		
	(0.07)			(0.07)		
Cooperative peers	0.62***	0.62***	0.64***	0.45***	0.46***	0.51***
	(0.13)	(0.13)	(0.12)	(0.14)	(0.14)	(0.13)
Non-cooperative peers				-0.15**	-0.14**	-0.12*
				(0.07)	(0.07)	(0.07)
OTHER*Cooperative peers	0.01			0.06		
	(0.19)			(0.19)		
OTHER*Non-cooperative peers				0.03		
				(0.10)		
Observations	148	74	74	296	148	148

Notes: OLS, robust standard errors in parentheses. *** denotes p<0.01, ** p<0.05 and * p<0.1. Panel A reports results together for subjects in the OBSERVING ONE PEER (1PO) and OBSERVING TWO PEERS (2PO) conditions. Panel B and Panel C report results separately for the 1PO and 2PO condition, respectively. OTHER is a dummy variable equal to one if the subject made a decision in the OTHER condition (Roma partner) and zero for choices in the SAME condition (Majority partner). See Table 2 for definitions of the variables "Cooperative peer(s)" and "Non-cooperative peer(s)". In Columns 1-3 the subjects in the NO PEERS OBSERVED (NPO) are excluded and the omitted group is "Non-cooperative peer(s)". In Columns 4-6 the omitted group is the NPO condition. In all Columns of all Panels, we control for gender and school grade.

Table 2.5: Discrimination in destructive and cooperative behavior, across social context

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A:								
Dependent variable	**Destructive choice in the Joy of Destruction Game**							
Sample	NPO	1PO and 2PO		1PO		2PO		INDIVIDUAL
	All	Observing destructive peer(s)	Observing non-destructive peer(s)	Observing destructive peer	Observing non-destructive peer	Observing destructive peers	Observing non-destructive peers	All
OTHER	-0.05	0.27***	-0.08	0.29**	-0.03	0.20	-0.11	0.01
	(0.08)	(0.09)	(0.06)	(0.11)	(0.09)	(0.13)	(0.08)	(0.06)
Observations	148	108	dssd	66	80	42	106	210
textbfPanel B:								
Dependent variable	**Cooperative choice in the Prisoner's Dilemma game**							
Sample	NPO	1PO and 2PO		1PO		2PO		INDIVIDUAL
	All	Observing cooperative peer(s)	Observing non-cooperative peer(s)	Observing cooperative peer	Observing non-cooperative peer	Observing cooperative peers	Observing non-cooperative peers	All
OTHER	0.01	-0.04	0.05	-0.07	0.04	0.05	0.05	0.01
	(0.07)	(0.13)	(0.05)	(0.16)	(0.07)	(0.19)	(0.07)	(0.06)
Observations	148	62	232	41	105	21	127	210

Notes: OLS, robust standard errors in parentheses. *** denotes p<0.01, ** p<0.05 and * p<0.1. The dependent variable in Panel A is a destructive choice in the Joy of Destruction game. In Panel B the dependent variable is a cooperative choice in the Prisoner's Dilemma game. OTHER is a dummy variable equal to one if the subject made a decision in the OTHER condition (Roma partner) and zero for choices in the SAME condition (Majority partner). Column 1 reports results for choices in the NO PEERS OBSERVED (NPO) condition. Columns 2-7 present results for the OBSERVING ONE PEER (1PO) and OBSERVING TWO PEERS (2PO) conditions. See Table 2 for definitions of the variables "Destructive peer(s)", "Non-destructive peer(s)", "Cooperative peer(s)" and "Non-cooperative peer(s)". Column 8 reports results for choices in the INDIVIDUAL condition. In all Columns of both Panels, we control for gender and school grade.

Table 2.6: Beliefs and conditional decisions

Sample	INDIVIDUAL		
	OTHER (1)	SAME (2)	P.p. difference (1)-(2); (p-value) (3)
Panel A: Destructive choice in the Joy of Destruction game			
Beliefs about destructive behavior of counterpart	0.49	0.35	13 (0.05)
Likelihood of destructive choice, conditional on counterpart being destructive	0.58	0.51	7 (0.33)
Likelihood of destructive choice, conditional on counterpart being non-destructive	0.33	0.28	6 (0.37)
Panel B: Cooperative choice in the Prisoner's Dilemma game			
Beliefs about cooperative behavior of counterpart	0.34	0.34	0 (1.00)
Likelihood of cooperative choice, conditional on counterpart being non-cooperative	0.09	0.16	-8 (0.09)
Likelihood of cooperative choice, conditional on counterpart being cooperative	0.30	0.28	3 (0.65)

Notes: Means. Beliefs and conditional decisions in the INDIVIDUAL condition. Panel A reports beliefs and conditional decisions in the Joy of Destruction game, while Panel B reports beliefs and conditional decisions in the Prisoner's Dilemma game. Conditional decisions were elicited using a strategy method, i.e. without knowing the action of the counterpart. OTHER and SAME indicate whether the subjects were deciding in the OTHER condition (Roma partner) or in the SAME condition (Majority partner). All differences are presented in percentage points and tested using a Chi-square test.

2.A Appendix 2

Table 2.7: OBSERVING TWO PEERS condition (2PO), mixed signal

	(1)	(2)	(3)	(4)	(5)
Panel A: Destructive choice in the Joy of Destruction game					
Sample	2PO				
	Two destructive peers	One destructive and one non-destructive peer	Two non-destructive peers	P.p. difference (1)-(3); (p-value)	P.p. difference (2)-(3); (p-value)
OTHER	0.88	0.4	0.09	79 (0.00)	31 (0.01)
SAME	0.67	0.62	0	67 (0.00)	62 (0.00)
P.p. difference OTHER-SAME; (p-value)	21 (0.10)	-22 (0.18)	9 (0.10)		
Observations	42	41	65		
Panel B: Cooperative choice in the Prisoner's Dilemma game					
Sample	2PO				
	Two cooperative peers	One cooperative and one non-cooperative peer	Two non-cooperative peers	P.p. difference (1)-(3); (p-value)	P.p. difference (2)-(3); (p-value)
OTHER	0.8	0.32	0.14	66 (0.00)	18 (0.10)
SAME	0.73	0.33	0.9	64 (0.00)	24 (0.02)
P.p. difference OTHER-SAME; (p-value)	7 (0.70)	2 (0.92)	5 (0.43)		
Observations	21	40	87		

Notes: Means. OBSERVING TWO PEERS (2PO) condition, more detailed classification of the signal from preceding players. Panel A reports prevalence of destructive behavior in the Joy of Destruction game. Panel B reports prevalence of cooperative behavior in the Prisoner's Dilemma game. OTHER and SAME indicate whether the subjects are deciding in the OTHER condition (Roma partner) or in the SAME condition (Majority partner). All differences are presented in percentage points and tested using a Chi-square test.

Table 2.8: The effect of peer behavior on the prevalence of destructive choices (Joy of Destruction game), by gender

Dependent variable	**Destructive choice in the Joy of Destruction game**					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Males						
Sample	1PO and 2PO			NPO, 1PO and 2PO		
	All	SAME	OTHER	All	SAME	OTHER
OTHER	-0.06			-0.17		
	(0.10)			(0.11)		
Destructive peer(s)	0.26**	0.27**	0.68***	0.01	0.01	0.54***
	(0.12)	(0.11)	(0.08)	(0.12)	(0.12)	(0.09)
Non-destructive peer(s)				-0.24**	-0.25**	-0.13
				(0.11)	(0.11)	(0.10)
OTHER*Destructive peer(s)	0.42***			0.53***		
	(0.14)			(0.15)		
OTHER*Non-destructive peer(s)				0.10		
				(0.15)		
Observations	142	71	71	226	113	113
Panel B: Females						
Sample	1PO and 2PO			NPO, 1PO and 2PO		
	All	SAME	OTHER	All	SAME	OTHER
OTHER	-0.09			0.09		
	(0.07)			(0.12)		
Destructive peer(s)	0.35***	0.35***	0.61***	0.20	0.20	0.26**
	(0.12)	(0.12)	(0.10)	(0.14)	(0.14)	(0.12)
Non-destructive peer(s)				-0.16	-0.16	-0.34***
				(0.10)	(0.10)	(0.10)
OTHER*Destructive peer(s)	0.25			0.06		
	(0.16)			(0.19)		
OTHER*Non-destructive peer(s)				-0.18		
				(0.15)		
Observations	152	76	76	216	108	108

Notes: OLS, robust standard errors in parentheses. *** denotes p<0.01, ** p<0.05 and * p<0.1. Panel A reports results for Male decision-makers and Panel B reports results for Female decision-makers. The results are reported together for subjects in the OBSERVING ONE PEER (1PO) and OBSERVING TWO PEERS (2PO) conditions. OTHER is a dummy variable equal to one if the subject made a decision in the OTHER condition (Roma partner) and zero for choices in the SAME condition (Majority partner). See Table 2 for definitions of the variables "Destructive peer(s)" and "Non-destructive peer(s)". In Columns 1-3 the subjects in the NO PEERS OBSERVED (NPO) are excluded and the omitted group is "Non-destructive peer(s)". In Columns 4-6 the omitted group is the NPO condition. In all Columns of both Panels, we control for school grade.

Table 2.9: Discrimination in destructive behavior, by gender

Dependent variable	**Destructive choice in the Joy of Destruction game**			
	(1)	(2)	(3)	(4)
Panel A: Males				
Sample	NPO	1PO and 2PO		INDIVIDUAL
		Destructive peer(s)	Non-destructive peer(s)	
OTHER	-0.17	0.36***	-0.06	0.00
	(0.11)	(0.10)	(0.10)	(0.09)
Observations	84	58	84	102
Panel B: Females				
Sample	NPO	1PO and 2PO		INDIVIDUAL
		Destructive peer(s)	Non-destructive peer(s)	
OTHER	0.09	0.14	-0.10	0.02
	(0.12)	(0.14)	(0.07)	(0.09)
Observations	64	50	102	108

Notes: OLS, robust standard errors in parentheses. *** denotes p<0.01, ** p<0.05 and * p<0.1. Panel A reports results for Male decision-makers and Panel B reports results for Female decision-makers. OTHER is a dummy variable equal to one if the subject made a decision in the OTHER condition (Roma partner) and zero for choices in the SAME condition (Majority partner). Column 1 reports results for choices in the NO PEERS OBSERVED (NPO) condition. Columns 2-3 present results for the OBSERVING ONE PEER (1PO) and OBSERVING TWO PEERS (2PO) conditions. See Table 2 for definitions of the variables "Destructive peer(s)" and "Non-destructive peer(s)". Column 4 reports results for choices in the INDIVIDUAL condition. In all Columns of both Panels, we control for school grade.

Table 2.10: The effect of peer behavior on the prevalence of destructive choices (Joy of Destruction game), by the education of parents

Dependent variable	**Destructive choice in the Joy of Destruction game**					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: At least one parent has a university degree						
Sample	1PO and 2PO			NPO, 1PO and 2PO		
	All	SAME	OTHER	All	SAME	OTHER
OTHER	-0.01			-0.17		
	(0.13)			(0.16)		
Destructive peer(s)	0.28	0.27	0.64***	0.04	-0.02	0.58***
	(0.18)	(0.18)	(0.14)	(0.19)	(0.19)	(0.16)
Non-destructive peer(s)				-0.23	-0.25*	-0.05
				(0.15)	(0.14)	(0.14)
OTHER*Destructive peer(s)	0.33			0.49*		
	(0.23)			(0.25)		
OTHER*Non-destructive peer(s)				0.16		
				(0.20)		
Observations	70	35	35	106	53	53
Panel B: None of the parents has a university degree						
Sample	1PO and 2PO			NPO, 1PO and 2PO		
	All	SAME	OTHER	All	SAME	OTHER
OTHER	-0.08			0.05		
	(0.08)			(0.12)		
Destructive peer(s)	0.37***	0.37***	0.63***	0.13	0.13	0.26**
	(0.11)	(0.11)	(0.09)	(0.12)	(0.12)	(0.11)
Non-destructive peer(s)				-0.23**	-0.23**	-0.37***
				(0.10)	(0.10)	(0.10)
OTHER*Destructive peer(s)	0.27*			0.13		
	(0.14)			(0.16)		
OTHER*Non-destructive peer(s)				-0.14		
				(0.14)		
Observations	162	81	81	238	119	119

Notes: OLS, robust standard errors in parentheses. *** denotes p<0.01, ** p<0.05 and * p<0.1. Panel A reports results for decision-makers who have at least one parent with a university degree and Panel B reports results for decision-makers who have both parents without a university degree (self-reported). The results are reported together for subjects in the OBSERVING ONE PEER (1PO) and OBSERVING TWO PEERS (2PO) conditions. OTHER is a dummy variable equal to one if the subject made a decision in the OTHER condition (Roma partner) and zero for choices in the SAME condition (Majority partner). See Table 2 for definitions of the variables "Destructive peer(s)" and "Non-destructive peer(s)." In Columns 1-3 the subjects in the NO PEERS OBSERVED (NPO) are excluded and the omitted group is "Non-destructive peer(s)". In Columns 4-6 the omitted group is the NPO condition. In all Columns of both Panels, we control for gender and school grade.

Table 2.11: Discrimination in destructive behavior, by the education of parents

Dependent variable	**Destructive choice in the Joy of Destruction game**			
	(1)	(2)	(3)	(4)
Panel A: At least one parent has a university degree				
Sample	NPO	1PO and 2PO		INDIVIDUAL
		Destructive peer(s)	Non-destructive peer(s)	
OTHER	-0.17	0.31	-0.01	-0.00
	(0.17)	(0.20)	(0.13)	(0.14)
Observations	36	23	47	48
Panel B: None of the parents has a university degree				
Sample	NPO	1PO and 2PO		INDIVIDUAL
		Destructive peer(s)	Non-destructive peer(s)	
OTHER	0.05	0.19*	-0.08	-0.04
	(0.12)	(0.11)	(0.08)	(0.09)
Observations	76	66	96	102

Notes: OLS, robust standard errors in parentheses. *** denotes p<0.01, ** p<0.05 and * p<0.1. Panel A reports results for decision-makers who have at least one parent with a university degree and Panel B reports results for decision-makers who have both parents without a university degree (self-reported). OTHER is a dummy variable equal to one if the subject made a decision in the OTHER condition (Roma partner) and zero for choices in the SAME condition (Majority partner). Column 1 reports results for choices in the NO PEERS OBSERVED (NPO) condition. Columns 2-3 present results for the OBSERVING ONE PEER (1PO) and OBSERVING TWO PEERS (2PO) conditions. See Table 2 for definitions of the variables "Destructive peer(s)" and "Non-destructive peer(s)". Column 4 reports results for choices in the INDIVIDUAL condition. In all Columns of both Panels, we control for gender and school grade.

Table 2.12: The effect of peer behavior on the prevalence of destructive choices (Joy of Destruction game), excluding observations where subjects did not answer all comprehension questions correctly

Dependent variable	**Destructive choice in the Joy of Destruction game**					
Sample	1PO and 2PO			NPO, 1PO and 2PO		
	All	SAME	OTHER	All	SAME	OTHER
	(1)	(2)	(3)	(4)	(5)	(6)
OTHER	-0.09			-0.04		
	(0.06)			(0.09)		
Destructive peer(s)	0.32***	0.33***	0.68***	0.09	0.09	0.40***
	(0.09)	(0.09)	(0.07)	(0.10)	(0.10)	(0.09)
Non-destructive peer(s)				-0.23***	-0.23***	-0.28***
				(0.08)	(0.08)	(0.08)
OTHER*Destructive peer(s)	0.37***			0.31**		
	(0.12)			(0.13)		
OTHER*Non-destructive peer(s)				-0.04		
				(0.11)		
Observations	233	119	114	354	179	175

Notes: OLS, robust standard errors in parentheses. *** denotes p<0.01, ** p<0.05 and * p<0.1. The effect of peer behavior on the likelihood of a destructive choice in the Joy of Destruction game, using the subsample of observations where subjects answered all control questions correctly. OTHER is a dummy variable equal to one if the subject made a decision in the OTHER condition (Roma partner) and zero for choices in the SAME condition (Majority partner). See Table 2 for definitions of the variables "Destructive peer(s)" and "Non-destructive peer(s)". In Columns 1-3 the subjects in the NO PEERS OBSERVED (NPO) are excluded and the omitted group is "Non-destructive peer(s)". In Columns 4-6 the omitted group is the NPO condition. In all Columns, we control for gender and school grade.

Table 2.13: Discrimination in destructive behavior (Joy of Destruction game), excluding observations where subjects did not answer all comprehension questions correctly

Dependent variable	**Destructive choice in the Joy of Destruction game**			
	NPO	1PO and 2PO		INDIVIDUAL
Sample		Destructive peer(s)	Non-destructive peer(s)	
	(1)	(2)	(3)	(4)
OTHER	-0.04	0.27***	-0.09	-0.01
	(0.09)	(0.10)	(0.06)	(0.07)
Observations	121	88	145	181

Notes: OLS, robust standard errors in parentheses. *** denotes p<0.01, ** p<0.05 and * p<0.1. Destructive behavior in the Joy of Destruction game, using the subsample of observations where subjects answered all control questions correctly. OTHER is a dummy variable equal to one if the subject made a decision in the OTHER condition (Roma partner) and zero for choices in the SAME condition (Majority partner). Column 1 reports results for choices in the NO PEERS OBSERVED (NPO) condition. Columns 2-3 present results for the OBSERVING ONE PEER (1PO) and OBSERVING TWO PEERS (2PO) conditions. See Table 2 for definitions of the variables "Destructive peer(s)" and "Non-destructive peer(s)". Column 4 reports results for choices in the INDIVIDUAL condition. In all Columns, we control for gender and school grade.

Table 2.14: The effect of peer behavior on the prevalence of destructive choices (Joy of Destruction game), controlling for order effects and experimenter fixed effects

Dependent variable	**Destructive choice in the Joy of Destruction game**					
Sample	1PO and 2PO			NPO, 1PO and 2PO		
	All	SAME	OTHER	All	SAME	OTHER
	(1)	(2)	(3)	(4)	(5)	(6)
OTHER	-0.08			-0.05		
	(0.06)			(0.08)		
Destructive peer(s)	0.26***	0.22**	0.61***	0.07	0.04	0.36***
	(0.08)	(0.09)	(0.07)	(0.09)	(0.09)	(0.08)
Non-destructive peer(s)				-0.19***	-0.17**	-0.21***
				(0.07)	(0.07)	(0.07)
OTHER*Destructive peer(s)	0.34***			0.31***		
	(0.11)			(0.12)		
OTHER*Non-destructive peer(s)				-0.02		
				(0.10)		
Observations	294	147	147	442	221	221

Notes: OLS, robust standard errors in parentheses. *** denotes p<0.01, ** p<0.05 and * p<0.1. OTHER is a dummy variable equal to one if the subject made a decision in the OTHER condition (Roma partner) and zero for choices in the SAME condition (Majority partner). See Table 2 for definitions of the variables "Destructive peer(s)" and "Non-destructive peer(s)". In Columns 1-3 the subjects in the NO PEERS OBSERVED (NPO) are excluded and the omitted group is "Non-destructive peer(s)". In Columns 4-6 the omitted group is the NPO condition. In all Columns, we control for a dummy variable indicating that the Prisoner's Dilemma game was played before the Joy of Destruction game, for a dummy variable indicating that the OTHER condition was introduced first, for experimenter fixed effects (five experimenters overall), and for gender and school grade.

Table 2.15: Discrimination in destructive behavior, controlling for order effects and experimenter fixed effects

Dependent variable	**Destructive choice in the Joy of Destruction game**			
	NPO	1PO and 2PO		INDIVIDUAL
Sample		Destructive peer(s)	Non-destructive peer(s)	
	(1)	(2)	(3)	(4)
OTHER	-0.05	0.32***	-0.04	0.01
	(0.08)	(0.08)	(0.06)	(0.06)
Observations	148	108	186	210

Notes: OLS, robust standard errors in parentheses. *** denotes p<0.01, ** p<0.05 and * p<0.1. OTHER is a dummy variable equal to one if the subject made a decision in the OTHER condition (Roma partner) and zero for choices in the SAME condition (Majority partner). Column 1 reports results for choices in the NO PEERS OBSERVED (NPO) condition. Columns 2-3 present results for the OBSERVING ONE PEER (1PO) and OBSERVING TWO PEERS (2PO) conditions. See Table 2 for definitions of the variables "Destructive peer(s)" and "Non-destructive peer(s)". Column 4 reports results for choices in the INDIVIDUAL condition. In all Columns, we control for a dummy variable indicating that the Prisoner's Dilemma game was played before the Joy of Destruction game, for a dummy variable indicating that the OTHER condition was introduced first, for experimenter fixed effects (five experimenters overall), and for gender and school grade.

Table 2.16: The effect of peer behavior on the prevalence of destructive choices (Joy of Destruction game), first implemented condition (between-subject design)

Dependent variable	**Destructive choice in the Joy of Destruction game**					
Sample	1PO and 2PO			NPO, 1PO and 2PO		
	All	SAME	OTHER	All	SAME	OTHER
	(1)	(2)	(3)	(4)	(5)	(6)
OTHER	-0.11			-0.11		
	(0.09)			(0.12)		
Destructive peer(s)	0.17	0.17	0.62***	0.05	0.06	0.50***
	(0.11)	(0.12)	(0.11)	(0.12)	(0.12)	(0.12)
Non-destructive peer(s)				-0.12	-0.10	-0.12
				(0.10)	(0.10)	(0.11)
OTHER*Destructive peer(s)	0.44***			0.45**		
	(0.15)			(0.17)		
OTHER*Non-destructive peer(s)				0.00		
				(0.15)		
Observations	147	83	64	221	125	96

Notes: OLS, robust standard errors in parentheses. *** denotes p<0.01, ** p<0.05 and * p<0.1. The effect of peer behavior on the likelihood of a destructive choice in the Joy of Destruction game, using only the first implemented condition (OTHER or SAME), thus mimicking a between-subject design. OTHER is a dummy variable equal to one if the subject made a decision in the OTHER condition (Roma partner) and zero for choices in the SAME condition (Majority partner). See Table 2 for definitions of the variables "Destructive peer(s)" and "Non-destructive peer(s)". In Columns 1-3 the subjects in the NO PEERS OBSERVED (NPO) are excluded and the omitted group is "Non-destructive peer(s)". In Columns 4-6 the omitted group is the NPO condition. In all Columns, we control for gender and school grade.

Table 2.17: Discrimination in destructive behavior, first implemented condition (between-subject design)

Dependent variable	**Destructive choice in the Joy of Destruction game**			
	NPO	1PO and 2PO		INDIVIDUAL
Sample		Destructive peer(s)	Non-destructive peer(s)	
	(1)	(2)	(3)	(4)
OTHER	-0.11	0.33**	-0.10	0.01
	(0.12)	(0.13)	(0.09)	(0.09)
Observations	74	47	100	105

Notes: OLS, robust standard errors in parentheses. *** denotes p<0.01, ** p<0.05 and * p<0.1. Destructive behavior in the Joy of Destruction Game, using only the first implemented condition (OTHER or SAME), thus mimicking a between-subject design. OTHER is a dummy variable equal to one if the subject made a decision in the OTHER condition (Roma partner) and zero for choices in the SAME condition (Majority partner). See Table 2 for definitions of the variables "Destructive peer(s)" and "Non-destructive peer(s)". In Columns 1-3 the subjects in the NO PEERS OBSERVED (NPO) are excluded and the omitted group is "Non-destructive peer(s)". In Columns 4-6 the omitted group is the NPO condition. In all Columns, we control for gender and school grade.

Chapter 3

How Stress Affects Performance and Competitiveness across Gender

Co-authored by Lubomír Cingl and Ian Levely

3.1 Introduction

It has been well established that men are on average more competitive than women and this phenomenon helps to explain gender differences in economic outcomes, including under-representation of women in certain industries and top-management positions (Gneezy, Niederle, and Rustichini 2003; Niederle and Vesterlund 2007; Niederle and Vesterlund 2011). The gender gap in willingness to compete has been found in western societies both in the laboratory and field experiments as well as in real markets (Flory, Leibbrandt, and List 2015; Jurajda and Munich 2011; Vincent 2013) and has been shown to predict real-life choices (Berge et al. 2015b; Buser, Niederle, and Oosterbeek 2014; Niederle and Vesterlund 2010).

Recent studies demonstrate that competitive situations are stressful (Buckert et al. 2015; Buser, Dreber, and Mollerstrom 2015; Fletcher, Major, and Davis 2008). By stress, we refer to a complex reaction that evolved in order to help organisms to deal with an uncontrollable threat to their major goal, like a threat to their physical or social survival (Dickerson and Kemeny 2004). Stress arises instinctively and imposes automatic behavioral effects on the decision maker (Starcke and Brand 2012) with potentially different

effects on men and women.[1]

Situations like university admissions, job interviews, asking for promotion and generally working in high-stakes environments involve heightened levels of acute[2] stress and occur in a competitive environment. Such situations are crucial events for many career paths and determine future economic outcomes. If women suffer more from the adverse behavioral effects of competition under stress than men, they may try to avoid such environments. If true, this fact could help to explain the observed gender gap in willingness to compete and the associated under-representation of women in highly competitive positions. In this paper we contribute to the literature by examining how acute stress affects performance and willingness to compete across gender.

This study is to our knowledge the first to examine the causal link between acute psychosocial[3] stress and competitive behavior. We employ a controlled laboratory economic experiment with 190 university students, 95 males and 95 females. A laboratory experiment allows for better control over confounding factors, such as selection bias into competitive and stressful situations. In our study, stress is exogenously introduced by a modified version of the Trier Social Stress Test for Groups (TSST-G; von Dawans, Kirschbaum, and Heinrichs 2011) where subjects go through either a stress treatment or a control condition in a between-subject design.

The experimental design we employ to study the change in performance and willingness to compete under stress is based on the paradigm of Niederle and Vesterlund (2007). Subjects in experimental sessions consisting of four men and four women are asked to

[1]Psychological research shows that women react differently to stressors (Kudielka and Kirschbaum 2005) and to different stressors than men (Stroud, Salovey, and Epel 2002). A recent theory posits that rather than "fight-or-flight" (Cannon 1932), women react in a "tend-and-befriend" manner, where tending means caring for self and offspring and befriending means affiliation with social groups to reduce general risk (Taylor 2006, Taylor et al. 2000).

[2]The stress response normally affects the organism in two ways: it supports the immediate coping strategies and suppresses the long-term processes that are not immediately necessary. Such a reaction is very demanding, therefore after some time the organism becomes exhausted, the supportive effects disappear while the long-term processes can be kept shut down, which may result in negative outcomes, including health and psychological problems. Hence, the behavioral effects of acute, short-term stress may starkly differ from chronic, persistent stress (McEwen 2012).

[3]We believe that currently the most common type of stress that people face in their lives is psychosocial stress, because it is the social status, not physical survival, that is being threatened in subjectively uncontrollable situations (Dickerson and Kemeny 2004). Stressors generally differ from each other by the effects they cause in the body: a physical stressor (stemming from, e.g. blood loss and sleep deprivation) may eventually produce a different response than a psychological stressor (e.g. interpersonal conflict or death in family) (Baum and Grunberg 1997, Clow 2001). In a related study of Buser, Dreber, and Mollerstrom (2015), no impact of stress on willingness to compete has been found which may be a result of the fact that the participants in their experiment were exposed to a physical stressor (putting a hand in ice-cold water).

individually perform by repeatedly adding up sets of four two-digit numbers within a time limit under varying payment schemes. A piece-rate incentive scheme is used first in a baseline condition and then repeated under the stress/control procedure to reveal the sole effect of stress on individual performance. A tournament incentive scheme follows where subjects' payoff depends on their performance relative to another randomly selected participant. Then the subjects choose their own payment scheme for the upcoming performance using any linear combination of the piece-rate and tournament payment scheme, which is our measure of willingness to compete (based on Gneezy and Pietrasz 2013). Subsequently they perform again and are rewarded accordingly. Additionally, using the same linear combination principle, subjects decide ex-post about the preferred payment scheme for their performance in the two piece-rate schemes which occurred before and after the exposure to the stress/control procedure. This allows us to be more specific regarding the underlying channels through which stress may affect the ex-ante willingness to compete. We are able to separate the effects of stress on general factors such as feedback aversion, risk aversion and confidence (which would also affect the ex-post decisions) from potentially worse (expected) performance under stress and from the preference for performing in a competitive environment under stress.

Our main finding is that stress reduces the ex-ante willingness to compete. For women, the decrease can be explained by worse performance in competitive environments under stress. The introduction of tournament incentives has a different effect on the performance of women in the stress treatment, compared to women in the control group. While the tournament incentives increase the average performance of women in the control group, the average performance of women in the stress treatment group drops, relative to the performance under piece-rate incentives. The associated confidence levels for tournaments are also lower for women in the treatment group, compared to the control.

Contrary to women, men's performance and confidence are not affected by stress. Men in both the control group and the stress treatment improve their score when tournament incentives are introduced and the effect is not statistically different across conditions. The gender difference in the effect of the stress treatment on tournament performance is not due to different compliance rates with respect to stress manipulation. Stress manipulation was successful for both genders as illustrated by a sharp increase in salivary cortisol levels following the TSST-G stress procedure. The physiological reaction is actually stronger for men, which is consistent with previous literature (Kajantie and Phillips 2006).

It is important to note that the willingness to compete drops only when the decision

is made before performing under stress (the ex-ante decision). When subjects do not have to subsequently perform under stress, but only state their willingness to compete for past performance conducted under piece-rate incentives (the ex-post decisions), we find no difference between the stress treatment and the control group, and this holds for both genders. This finding suggests that preferences for competition were not affected by stress. For women, the lower ex-ante willingness to compete seems to be connected to the expected decline in performance in tournaments under stress and also worse related confidence. For men, there is no effect of stress on performance in tournaments or confidence. Rather, the comparison of the ex-ante and ex-post willingness to compete suggests that the change in the ex-ante willingness to compete under stress is preference-based—men seem to have a lower willingness to go through a competitive environment under stress.

Overall, our findings suggest that women may be disadvantaged when required to compete in a stressful setting, and have broader implications for understanding how men and women approach competition. While Niederle and Vesterlund (2007) find that tournament incentives led to higher performance among both men and women, other studies (Gneezy, Niederle, and Rustichini 2003; Gneezy and Rustichini 2004) find that while men increase output in response to competitive incentives, this is not the case for women. These results may be caused by the relationship revealed in our study: potentially some of the tournaments were stressful for the subjects and this is why the competitive incentives did not improve the performance of women. Our study also contributes to the discussion on the sources of persistent under-representation of women in high-stakes positions and industries, such as in leadership positions in politics and business (Bertrand and Hallock 2001; Bertrand 2009). These are usually environments that are both highly competitive and stressful. If women know they do not perform well under these types of environments, they may decide to stay out.

3.2 Experimental design

All subjects completed several incentivized tasks which measure performance under piece-rate and tournament incentive schemes and willingness to compete. Our experimental manipulation consists of two treatments applied between-subjects, one in which subjects were exposed to a psychosocial stressor in the form of a TSST-G treatment procedure (*stress treatment*) and a *control* treatment.

3.2.1 Experimental tasks

We measure competitiveness using a design based on Niederle and Vesterlund (2007) and Gneezy and Pietrasz (2013). Subjects completed a counting task, twice under a noncompetitive piece-rate scheme, then under a competitive tournament scheme, and then were asked which combination of these compensation schemes they preferred for the next counting round.

The counting task required the subjects to add up series of four two-digit numbers. In each particular task two-minutes were allotted and subjects were given the chance to solve as many of these addition problems as they were able to. Subjects had a trial round to get themselves familiar with the task. While participants received immediate feedback on the number of correct answers in the particular task, they were not informed of others' results. Correct results in the counting task were incentivized according to different compensation schemes, which will be explained next.

Under the *piece-rate* compensation scheme, participants earned CZK 25 (about EUR 1) per each correct answer. Performance under the piece-rate scheme serves as a baseline measure of ability and effort in the counting task. Subjects performed twice under the piece rate compensation scheme: once before the *stress treatment/control* procedure (Task 1, *Piece rate before treatment*) and once after the *stress treatment/control* procedure (Task 2, *Piece rate under treatment*). Comparing Task 1 and Task 2 therefore allows us to directly measure the effect of the stress treatment on performance within subjects.

In Task 3, *Tournament under treatment*, correct answers were rewarded according to the *tournament* compensation scheme: each participant was informed that s/he would be randomly matched with another participant in the room (there were always four males and four females present) and that whoever had the most correct answers would receive CZK 50 per correct answer and the participant with fewer correct answers would receive CZK 0 per correct answer. In case of a tie, each participant received CZK 25 per correct answer, as in the piece-rate scheme.

In Task 4, *Choice of compensation scheme for future performance*, subjects chose ex-ante how they would be compensated for their performance in the counting portion of Task 4. They did so by splitting 100 points between the *tournament* and the *piece-rate* compensation schemes, as in Gneezy and Pietrasz (2013). For each point invested in the *piece-rate* scheme, they earned CZK 0.25 per correct answer in the subsequent task. For each point invested into the *tournament* compensation scheme, they earned CZK 0.5

per correct answer, but only if they had more correct answers in Task 4 than another randomly selected participant in Task 3, and received nothing per each point invested in the tournament scheme if they answered fewer questions. In the case of a tie, each point invested in the tournament account was rewarded according to the piece-rate scheme (CZK 0.25 per answer). Thus, if subjects invested all points into the *piece-rate* scheme, they were paid CZK 25 per correct answer, as in Task 1 and Task 2. If all points were invested in the *tournament* scheme, they received CZK 50 per question if they answered more questions, 0 if they answered fewer and 25 in case of a tie, as in Task 3. If they invested some points in the *tournament* scheme and some in the *piece-rate* scheme, they were paid according to a linear combination of the two compensation schemes.

We should emphasize that the choice of compensation scheme in Task 4 cannot be driven by prosocial concerns or expectations regarding who self-selects into the tournament, as a subject's performance in Task 4 was always compared to Task 3 performance of another randomly selected subject and this information was highlighted in the instructions. Therefore, subjects knew that their decision to enter the tournament did not have payoff consequences for anyone else.

The choice of compensation scheme in Task 4 is our main measure of the willingness to compete. To estimate the causal effect of stress on the willingness to compete, we compare the share of the 100 points invested in the *tournament* in Task 4 by subjects in the *stress treatment* and *control*. To determine whether potential treatment differences in the willingness to compete are influenced by factors such as confidence, risk preferences, feedback aversion, or rather by preferences for performing in a competitive environment (Niederle and Vesterlund 2007), we implemented two additional tasks, in which subjects competed based on past performance, rather than subsequently performing the counting task.

In Task 5, *Choice of compensation scheme for past performance before treatment*, subjects again split 100 points between the *tournament* and *piece-rate* schemes, but were paid according to performance in Task 1, which took place before the stress manipulation was introduced. Before Task 5, subjects were reminded that Task 1 was incentivized with a piece-rate scheme and that it took place in the first room, indicating that it took place before the stress/control procedure. Additionally, they were reminded how many problems they solved correctly in Task 1. The Task 5 decision should therefore capture willingness to compete, but, since the decision is ex-post for performance which occurred outside the stress treatment, without taking into account preferences for performing under

competition or (beliefs about) the potential negative effect of stress on performance.

In Task 6, *Choice of compensation scheme for past performance under treatment*, subjects also split 100 points between the *tournament* and *piece-rate* schemes, but were paid according to performance in Task 2, which took place after the *stress/control* manipulation was introduced. Instructions for Task 6 reminded subjects about the timing of Task 2 and their performance. Therefore, if stress negatively impacts performance, and thus possibly changes subjective beliefs about relative performance, this should influence the subjects' decisions in both Task 4 and Task 6. The additional motives present in Task 4 Choice compared to Task 6 Choice are only the preference for performing in a competitive environment, plus (beliefs about) performance in tournaments under *stress/control* condition.

As additional ways of estimating the role of confidence in competitiveness decisions, subjects were given non-incentivized questions regarding their perceived rank among all eight participants in the given session. Specifically, they were asked about their perceived rank when performing under the *piece-rate* scheme outside the *stress/control* treatment (Task 1), under *piece-rate* scheme under the *stress/control* treatment (Task 2) and under *tournament* under the *stress/control* treatment (Task 3). Finally, we conducted Task 7 to measure risk preferences using a setting based on Dohmen et al. 2010. In this task, subjects were asked to repeatedly choose between a lottery, which was always kept the same at CZK 240 versus CZK 0 with 50% probability each, and a safe payment, which was gradually increasing from CZK 0 to CZK 240 in steps of CZK 20.

In order to limit possible hedging, subjects were informed that two out of the seven tasks (Task 1-Task 7) would be randomly selected for payment at the end of the experiment. Full experimental instructions for Tasks 1-7 are available upon request.

3.2.2 Experimental manipulations

We experimentally induce stress in the laboratory, using a modified version of the TSST-G (Kirschbaum, Pirke, and Hellhammer 1993; von Dawans, Kirschbaum, and Heinrichs 2011). This procedure was designed to induce mild psychosocial stress in the *stress treatment* group, along with a *control* procedure designed to similarly prime subjects yet to keep stress levels constant. The TSST-G has been shown to be the most efficient experimental method of elevating levels of cortisol, a hormone associated with stress (Dickerson and Kemeny 2004).

The standard *stress treatment* protocol consists of two parts: a public speaking task and a mental arithmetic task. Both tasks are performed by subjects one-by-one in front of a "committee" (2 experimenters), who sit at a table in front of the participants wearing white lab coats and are instructed not to give any feedback and to maintain a neutral facial expression throughout the procedure. The procedure is recorded by a video camera that is prominently visible. These steps are intended to increase the stressfulness of the procedure. The setting of the room is depicted in Appendix Figure 3.6.

We slightly modified the standard *stress treatment* protocol in several ways. Firstly, subjects were separated by dividers and wore headphones with ambient traffic noise during the entire TSST-G procedure, except when speaking to the committee. This was done to prevent subjects from hearing the speeches of others and potentially developing subjective rankings in ability. Secondly, we modified the public speaking task both to avoid deception and priming effects.[4] In our version of the public speaking task, participants were told to imagine a situation in which they had been caught cheating during an important academic examination and should defend themselves in front of a disciplinary committee. Subjects were instructed that they should do their best, despite the fact that this was a mock defence. As in the original protocol, this set-up required participants to talk extensively about their personal qualities. Subjects were interrupted if talking fluently for too long and were asked additional questions. Thirdly, in the second portion of our modified TSST-G procedure, subjects in the *stress treatment* were again called individually and asked to recite the alphabet backwards in steps of two, starting from a given letter. For example, if given a letter Z, they should recite Z, X, V,...[5] Subjects had to recite for a minute and were corrected if a mistake was made.

The *control* procedure generally exposed subjects to similar conditions, both cognitively and physically, but with minimal stress attributes. Subjects were asked to read an article about academic dishonesty, silently for the first fourteen minutes and then aloud

[4]The modifications with respect to deception concerned mainly the information given to the participants in the stress treatment; they were not told that the panel members were trained in behavioral analysis, or that the video recordings would later be analyzed as is the case in the original TSST-G script in von Dawans, Kirschbaum, and Heinrichs (2011). We were also concerned about possible priming effects, since the original procedure is framed as a job-interview, which could have influenced competitiveness and performance on the main task independently of the stress reaction. This is why we modified the framing of the speaking task.

[5]This is in contrast to the standard TSST-G, in which subjects are asked to count backwards (numerically) in various steps. We amended this portion of the procedure to avoid a confound with the counting portion of the competition task (i.e. stress levels may have been correlated with ability in the counting portion of the Competitiveness experiment).

for two minutes.[6] In the second part of the procedure, they collectively recited the alphabet out loud for a minute. Two experimenters were again present in the room in the control procedure, but wore normal clothes and behaved naturally. The subjects in the *control* group also wore headphones with ambient noise and were separated with cardboard dividers, to mimic conditions in the *stress treatment* group. The complete version of our modified TSST-G *stress treatment/control* protocol is available upon request.

The exact timing of the *stress/control procedure* with respect to Tasks 1-7 is summarized in Figure 3.1. After arriving at the laboratory and initial procedures, the instructions for the counting task were read by the experimenter, and then subjects completed a trial round. Next, subjects completed Task 1 – *Piece rate before treatment* in the laboratory. After Task 1, the first saliva sample was collected, and subjects filled out the first part of the multidimensional mood questionnaire (MDMQ, Steyer et al. 1997). The instructions for the *stress* or the *control* procedures of the TSST-G protocol were then handed out and read aloud by the experimenter (each session included only one treatment group). The subjects had two minutes for preparation, after which they were taken to a neighboring room for the *stress/control* procedure. The remaining tasks of the experiment (Task 2-7) were carried out in this space, so that the decisions were truly made in a *stress/control* environment. Subjects were solving the tasks at a PC that was located directly adjacent to the space where they stood for the *stress/control* procedure.

Subjects then completed the first portion of the *stress/control* procedure—either the public speaking task in the *stress treatment* group, or the reading task in the *control*. Immediately after the *stress/control* procedure, subjects sat down and completed Task 2 – *Piece rate under treatment* and Task 3 – *Tournament under treatment* on their computers. They were asked to stand up when finished and wait for others. After this, the second part of the *stress/control* procedure was carried out (the alphabet task). Immediately afterwards, subjects were asked to sit at their computers, to provide the second saliva sample and then continue with Tasks 4-6 (choices of compensation scheme for future and past performances), with the confidence questions and with Task 7 – *Risk preferences* measure. Subjects left the *stress/control* procedure room after completing Task 7 and returned to the laboratory, where the third saliva sample was collected.

[6]The timing of the control procedure is chosen to mimic the activities of the last-speaking subject in the stress procedure.

3.2.3 Sample and procedures

The experiment was carried out at the Laboratory of Experimental Economics in Prague in February 2014 and May 2015, with 24 sessions in total. Subjects were recruited using a standard recruitment database, ORSEE (Greiner 2004). The final sample is composed of 190 subjects, 95 males and 95 females, who are primarily undergraduate students (82%), majoring mostly in in economics, business and related fields (61%).[7] Subjects signed an informed consent form once they arrived to the laboratory, emphasizing that they were free to leave during any part of the experiment, an option which only one subject decided to take. The study was approved by the Internal Review Board of the Laboratory of Experimental Economics.

Each session consisted of eight subjects, four males and four females. The gender composition was not directly mentioned in any way (following Niederle and Vesterlund 2007), but the seating plan in the laboratory was set in a way that it was easily observable.[8] Each session included only the stress treatment or control group, for logistical reasons. The order of sessions by treatment was randomized, balancing the day of the week and time of the day. To avoid the intra-day variability of cortisol concentration all sessions were performed after 3PM. The experiment was conducted in Czech and sessions were administered by one experimenter (male), one assistant (female) and two separate "committee" members for the TSST-G procedure (a male and a female). The average length of the experiment was slightly less than 2 hours and the average payout was CZK 516.

For recruitment, we announced a two-hour experiment with an expected payment of CZK 500 (around EUR 20) including a guaranteed show-up fee of CZK 150. No particular information about the nature of the experiment was mentioned in the invitation email, which may have influenced self-selection into the experiment. The subjects were only given instructions to abstain from eating, smoking and intaking any medical substances prior to the experiment, which was done to avoid distorting cortisol measurement. After registering for the sessions, subjects filled in an on-line questionnaire that included the trait anxiety inventory (Spielberger et al. 1983), questions on risk-taking behavior in dif-

[7] In total 192 subjects participated in the experiments. However, one female subject is dropped from the analysis because she decided to leave during the experiment and one male subject is dropped because he was recruited as a last-minute replacement for a subject who did not show-up, and the replacement subject did not meet the selection criteria. The other subjects were unaware of this.

[8] In a questionnaire at the end of the experiment, the majority of subjects correctly reported the number of males in the group (74%) or the share of males in the group (80%).

ferent domains (after Dohmen et al. 2010), the BFI personality inventory (Rammstedt and John 2007), and questions on personal behavior that would distort the measurement of cortisol (smoking, medication intake, psychiatrist/psychologist treatment or any disorder of this type). These questionnaires were completed two days prior to the experiment. Compliance with instructions from the invitation email was checked once subjects arrived for the experiment.

Upon arriving at the laboratory, subjects read and signed an informed consent form, were fitted with heart-rate monitors[9] and completed the "Big 5" personality questionnaire (Goldberg, 2010). In the main part of the experiment, Tasks 1-7 were carried out in the laboratory and in the adjacent room, where the stress/control procedure took place, as described in the previous subsection. After completing all seven experimental tasks, the subjects returned to the laboratory, filled in the second part of the MDMQ questionnaire and performed a standard D2 attention test (Brickenkamp and Zillmer 1998). A short questionnaire on personal characteristics followed. At the end of the experiment, two tasks were randomly selected for payment in front of all participants, the subjects were paid in private and dismissed. Subjects in the stress treatment group went through a careful debriefing of the TSST-G procedure before they left.

3.3 Results

Willingness to compete

We begin by analyzing investment in the tournament payment scheme in Task 4, which serves as our baseline measure of willingness to compete. This decision captures both preferences for competitive outcomes as well as preferences for engaging in a competitive activity and expectations of one's future performance under competition. Recall that in Task 4 subjects allocated 100 points between a tournament and piece-rate incentive scheme before completing the counting portion of the task. The results from Task 4 are presented in Figure 3.2 and panel A of Table 3.1. Overall, subjects allocated slightly less than half of their allocation, 46.68 points, into the tournament incentive scheme. We find that stress does indeed affect competitiveness: subjects in the stress treatment invested 7.72 fewer points in the tournament scheme than subjects in the control group, which is statistically significant according to a rank-sum test ($p = 0.046$).

[9]Polar RS400, Polar Electro, Finland.

We confirm this result by regressing the points invested into the tournament scheme in Task 4 on a dummy equal to 1 if the subject was assigned to the *stress treatment*. We control for gender as well as baseline performance in Task 1 (i.e. before the treatment intervention) and cluster standard errors at the session level.[10] As reported in column 1 of Table 3.2 we find that the stress treatment was associated with investing 7.59 fewer points in the tournament scheme ($p = 0.024$).[11]

Consistent with the literature, we also find that gender has a strong influence on choices in Task 4, with women investing 25.27 fewer points in the tournament investment scheme than men (rank-sum test, $p = 0.00$), as reported in Figure 3.2 and panel A of Table 3.1. This is also confirmed by the regression results in column 1 of Table 3.2, in which we observe that women invested an average of 22.06 fewer points, after controlling for treatment and baseline performance, ($p = 0.00$).

The stress treatment has a similar effect on men and women, with respect to willingness to compete for future performance. In Figure 3.2 and panel A of Table 3.1, we see that the lower investment in the tournament payment scheme in Task 4 in the stress treatment that we observe on average holds for both male and female sub-samples, though the treatment differences are separately not statistically significant, due to the smaller sample size. In column 2 of Table 3.2, we add an interaction term between the female and stress treatment dummies to the regression on points invested into the tournament in Task 4 and observe no statistically significant gender difference ($p = 0.926$). In columns 3-4, we run regressions separately on the male and female sub-samples and find that the coefficients for the stress treatment are virtually identical, though both coefficients are marginally insignificant: $p = 0.123$ and $p = 0.124$ for the male and female sub-samples, respectively.

Performance and competitive incentives

We now turn to the counting portions of the tasks in order to examine how stress and gender affect performance, and whether this differs between competitive and non-competitive

[10] We use clustering to account for the correlation between outcomes within sessions that may arise due to shared experiences within the session, such as observing other subjects before the experiment. This results in 24 clusters. We confirm that the small number of clusters does not affect results by running a robustness check using the wild bootstrap method. Results are available upon request.

[11] Since most studies use binary measures to measure willingness to compete, we perform a robustness test in which we classify subjects as competitive if they invest more than 50/100 points into the tournament incentive scheme in Task 4 and estimate the effects of the stress treatment and gender using a probit model. Results are similar to the linear measure. See Appendix Table 3.5

settings. Understanding the effect of stress on performance is essential to our research question in its own right, but may also help to identify a potential channel through which stress affects willingness to compete. While subjects received no feedback on the performance of others in the counting portions of any of the tasks, they may nonetheless have inferred that their chances of winning the tournament were higher or lower depending on their performance in general and their perceived relative performance across the different tasks. The stress treatment could influence willingness to compete through performance, either by objectively affecting the number of correctly answered questions or subjectively through beliefs about relative performance.

Results from performance in the counting portions of Tasks 1-4 are presented in the upper panel of Figure 3.3 and Table 3.3. In panel A of Table 3.3, we see that there is virtually no difference in the number of correctly answered problems between the treatment and control groups under the piece-rate incentive scheme in Task 1 (rank-sum, $p = 0.931$). Since Task 1 was completed before the stress treatment was implemented, the lack of a significant difference here simply indicates that our randomization of treatment groups was successful. This holds for both the male and female sub-samples independently.

Performance in the counting portion of Task 2 captures the effect of the stress treatment on performance under the piece-rate incentive scheme. While on average the stress treatment group correctly answered 0.19 fewer problems than the control group (6.37 vs. 6.56 problems), the difference is not statistically significant (rank-sum test, $p = 0.560$). As before, this result holds for both the male and female subsamples, independently.

Recall that performance in Task 3 is influenced by both the stress treatment as well as the tournament payment scheme. Here, in contrast to Tasks 1 and 2, we see a significant treatment effect, with performance falling among the stress group, who answered only 6.24 problems correctly, compared to 7.14 in the control group (rank-sum, $p = 0.018$). This difference is much larger among females than males: women in the stress-treatment answered 1.37 fewer questions on average than those in the control group (rank-sum, $p = 0.003$), while the corresponding treatment difference for men is less than one third the size, 0.41 points, and is not statistically significant (rank-sum, $p = 0.562$).

We confirm this pattern using regression analysis, which is presented in Table 3.4. We regress performance under tournament incentives in Task 3 on the stress treatment, a dummy equal to 1 if the subject is female and baseline performance in Task 1, with standard errors clustered at the session level. We find that the stress treatment lowers

performance by 0.84 correctly answered questions on average ($p < 0.001$).[12] In column 2, we add a *stress treatment*female* interaction, and the results indicate that the effect of the stress treatment is driven by the female sub-sample and is gender-specific ($p = 0.019$). In columns 3-4, we estimate the effects separately for males and females: the stress treatment lowers female subjects' performance by 1.45 questions ($p < 0.001$), while the coefficient for male subjects does not differ statistically from zero ($p = 0.513$).

We can also observe the effect of tournament incentives on performance by examining the difference between the number of problems correctly solved in Tasks 3 and 2, and comparing this result across treatments and gender.

The overall trend can be observed in the bottom panel of Figure 3.3, which graphs the difference in the number of correctly solved problems in Tasks 3 and 2. Overall, subjects in the control group correctly answered 0.57 more questions in Task 3 under the competitive compensation scheme (t-test, $p = 0.002$). This is true for both for men and women in the control group, who answered 0.42 (t-test, $p = 0.091$) and 0.73 (t-test, $p = 0.007$) more questions correctly in Task 3 than in Task 2, respectively. For men in the stress treatment, performance is only slightly better in the competitive incentive scheme, with a difference of 0.23 correctly answered questions, which is not statistically different from zero (t-test, $p = 0.386$). For female subjects in the stress treatment, however, we see a different trend: performance in the tournament incentive scheme actually falls by 0.49 correct problems on average (t-test, $p = 0.096$).

In column 5 of Table 3.4, we regress the difference between correctly answered problems in Tasks 3 and 2, which can be interpreted as the effect of the tournament incentive scheme on performance, on stress treatment and gender. The results indicate that the stress treatment diminishes the effect of the tournament incentive scheme by 0.70 questions on average ($p = 0.008$). As before, in columns 6-8 we see that this is driven by the female sub-sample, while there is no statistically significant effect for men ($p = 0.586$).

Overall, these results indicate that female subjects perform significantly worse in a competitive setting when exposed to the stressor. Interestingly, women do not perform significantly worse when under stress during the piece-rate compensation scheme (Task 2), nor do women in the control group perform worse under the tournament incentive scheme (they actually perform significantly better under tournament incentives). Rather, it seems that it is the combination of stress and competition that decreases women's

[12] We run similar regressions on performance in Tasks 1 and 2, which confirm the lack of treatment effect we observe through rank-sum tests. Results are presented in Appendix Table 3.6

performance.[13] We do not find any such pattern for men, whose performance under tournaments is not significantly affected by the stress treatment.

Ex-post willingness to compete

The decisions in Tasks 5 and 6 give further insight into the mechanism behind the change in competitiveness in Task 4, which measured the willingness to compete for future performance.

In Task 5, subjects decided how much to invest in the tournament payment scheme ex-post, and the result of the tournament was decided by their performance in the counting Task 1 (i.e. under the piece-rate payment scheme and before the stress/control treatment). In contrast to the ex-ante competition decision in Task 4, we do not find a significant difference between the control and treatment groups for investment in the tournament in Task 5. On average, subjects in the control group invested 40.19 versus 41.20 in the stress treatment group, rank-sum $p = 0.826$ (panel A of Figure 3.4 and Panel B of Table 3.1). We fail to find a statistically significant treatment difference for either men or women as well.

In Task 6, subjects made an ex-post decision regarding their performance in Task 2 (piece-rate, under stress/control treatment). As in Task 5, we fail to find a statistically significant difference in the willingness to compete between treatments. Subjects in the control group invested 41.14 points into the tournament, while those in the treatment group invested 39.64 points on average, and the difference is not significant according to the rank-sum test, $p = 0.70$. Results are presented in panel C Table 3.1 and Panel B of Figure 3.4.[14]

Since Task 2 was completed after the stress treatment, changes in performance or perceived relative performance in response to the stressor should affect competitiveness in Tasks 4 and 6 similarly. The lack of result in Task 6 would thus suggest that the difference in competitiveness we see in Task 4 is not caused by a difference in perceived ability as a result of the stress treatment alone. Together, the results from Tasks 5 and 6 suggest that the decrease in competitiveness that we see in the stress treatment in Task

[13] We also consider performance in Task 4, though interpretation is less clear, since the incentive scheme is endogenous. On average, the stress treatment does not have a statistically significant effect on the number of correctly answered questions in Task 4 (rank-sum test, p=0.650). However, women in the stress treatment correctly completed 0.5 fewer correct problems (rank-sum, p=0.09). Regression results confirm this; see Appendix Table 3.7.

[14] We present regression results for investment in the tournament incentive scheme in Tasks 5 and 6 in Appendix Table 3.8. We do not find any significant treatment effects for either men or women.

4 is related to completing the task both under stress and in a competitive setting, rather than either element alone.

The gender difference in competitiveness that we observe in both treatments in Task 4 holds in both Tasks 5 and 6 as well, as visible from Figure 3.4.

Physiological stress response

Thus far, we have assumed that our version of the Trier-Social Stress test was successful in producing a stress response in subjects. To confirm this, we now turn to the cortisol levels of subjects throughout the experiment. Results are presented in Figure 3.5. As expected, there is no statistically significant difference in cortisol levels between treatment groups for the first cortisol sample, taken before the treatment intervention (t-test, p=0.51). Subsequent cortisol levels, taken after the second portion of our modified version of the TSST-G and after Task 7, respectively, are significantly higher for the stress treatment group than for the control group (cortisol samples 2 and 3, t-test, p=0.00). While cortisol levels after both the first and second rounds of the TSST-G procedure actually decrease slightly for the control group, there is a large and statistically significant increase in cortisol levels for those in the stress treatment group for the second and third samples (t-test, p=0.00).[15]

3.4 Discussion

We have presented results demonstrating that stress lowers willingness to compete. In this section, we discuss potential channels through which the treatment might produce this effect and, particularly for women, how this may be related to the lower performance under tournament incentives that we observe for female subjects in the *stress-treatment* group. The first mechanism that we consider is a change in preferences under heightened stress. Our design allows us to distinguish between two types of preference related to willingness to compete: preferences for engaging in a competitive activity (i.e. performing in the counting task with tournament incentives) and preferences for competitive outcomes. Since we find a treatment difference only in Task 4, when the competition decision was made *before* subjects completed the counting portion of the task, but not in the ex-post

[15] In appendix Table 3.9 we estimate the average treatment effect on the treated by using stress treatment as an instrument for cortisol levels and estimate the effects on willingness to compete in Task 4: the results are robust.

decisions in Tasks 5 and 6, our results seem to rule out the latter. This is in contrast to gender differences in willingness to compete. Consistent with findings in Niederle and Vesterlund (2007), we find that women are less competitive across all three investment decisions, which suggests different preferences for competitive outcomes.

Preferences for engaging in competition are closely linked with performance, and for women, we find that the *stress treatment* is associated with a decrease in performance under tournament incentives in Task 3. This likely indicates one of two closely related underlying effects (or a combination thereof): stress may affect preferences for engaging in competition, which may in turn lower effort, and consequently performance, or stress may lower the ability of women under tournament incentives. Unfortunately, we have no way of disentangling these two potential effects, since we cannot reliably measure effort independent of performance.

Regardless, our results suggest that, for women, the *stress treatment* lowers willingness to compete through performance. Even though subjects were unaware of the number of questions that others correctly answered, they observed their own performance under both the piece-rate and tournament compensation schemes, and likely based their investment in the tournament on their relative performance in these rounds.

To test this, we conduct a sensitivity analysis, which is reported in Appendix Table 3.10. We regress willingness to compete in Task 4 on the *stress treatment* dummy, then add performance in Task 3 as an additional explanatory variable. As predicted, Task 3 performance is positively correlated with the amount invested in the competition in Task 4.[16] The addition of Task 3 performance to the regression model lowers the *stress treatment* coefficient and increases the R-squared of the model. When we drop the *stress treatment* dummy from the model, the coefficient for Task 3 performance remains virtually unchanged. The R-squared is identical across the two models; after controlling for Task 3 performance, the stress treatment adds no explanatory power. This strongly suggests that, at least for women, the *stress treatment* affects willingness to compete in Task 4 principally by affecting performance—whether it is through preferences for engaging in competition and resulting effort levels or through ability.

It is also plausible that stress affects subjective beliefs about performance. To this

[16]Since our previous results indicate that the stress treatment is causally linked to women's performance in Task 3, these two variables are endogenous in the model, and therefore one must interpret the coefficients and standard errors with caution. However, comparing coefficients, standard errors and R-squared values across models nonetheless provides insight into the channel through which the *stress treatment* operates.

end, we measured subjective confidence for each round, after subjects had completed the entire experiment, by eliciting beliefs about the subject's rank among the 8 subjects in the session for each of the counting tasks (1-3). In Appendix Table 3.11, we regress confidence in Task 3 on treatment and observe that stress lowers confidence in tournaments, though only among women.[17] When we add Task 3 performance to the model, however, the treatment variable is no longer significant. When the treatment variable is dropped, the R-squared and standard errors remain the same. This suggests that the *stress treatment* does not affect subjective confidence levels, but rather affects confidence by objectively lowering women's performance in tournaments. For men, we find no treatment effect on confidence under tournament incentives. Importantly, we also do not find any effect of the stress treatment on confidence for tasks completed under piece-rate incentives before and under treatment (Task 1 and Task 2, respectively, see Appendix Figure 3.7), which is consistent with our non-result regarding the ex-post willingness to compete.[18]

Another possibility is that stress influences competitiveness through risk preferences. Cahlíková and Cingl (2016) find that a similar version of the TSST-G leads to higher levels of risk aversion, especially for men. Since the tournament incentive scheme increases subjects' exposure to risk, greater risk aversion might lead to lower willingness to compete. However, a change in risk-preferences would also affect ex-post willingness to compete, and we do not observe any effect in Tasks 5-6. In our sample, moreover, we fail to find any significant relationship between the *stress treatment*, and risk preferences elicited in Task 7. In fact, those in the *stress treatment* actually had slightly higher certainty equivalents than those in the control group, on average, though the difference is not statistically significant (rank-sum, p=0.334). This is consistent within both the male and female subsamples, (rank-sum, p=0.501 and p=0.698, respectively).[19] These results suggest that

[17] Again, one would expect confidence to be confounded with treatment, as well as performance, and adding both to the right-hand side creates an endogeneity problem, though we can still draw inferences by comparing the models.

[18] Goette et al. (2015) find that confidence under stress (with respect to past performance outside stress) differs across high-anxiety and low-anxiety individuals. We run a robustness check with respect to this possibility. Our baseline measure of anxiety was elicited two days prior to the experiment using the State-Trait Anxiety Inventory (STAI; Spielberger et al. 1983). Dividing our sample into high-anxiety and low-anxiety individuals using a median split, we find that our confidence results are robust: the negative effect of stress on confidence under tournament incentives holds for both high- and low-anxiety women and we do not find any significant effect of stress on confidence for tasks completed under piece-rate incentives. Results available upon request.

[19] We randomly chose two of the seven tasks (1-7) for payment, and thus it is possible that decisions in the risk task were affected by decisions in previous rounds. As stress lowered willingness to compete, leading to lower risk exposure in Task 3, this may have caused subjects in the stress treatment to make riskier decisions in Task 7, independent of risk preferences. Therefore, our measure of risk preferences

risk preferences are not a mechanism by which stress affects willingness to compete in our sample.

Using the d2 attention test, we further check that our results are not driven by cognitive load. The stress treatment and control group do not differ in terms of speed in the attention test (a rank-sum test, $p = 0.975$), total number of mistakes ($p = 0.312$), percentage of mistakes ($p = 0.174$) and error-corrected performance ($p = 0.679$). These results hold also for the male and female sub-samples separately.

Moreover, subjects were asked to rate their understanding of the experimental instructions in the end-questionnaire and we see that comprehension was high. Using a scale from 0 (not clear) - 10 (completely clear), the mean score is 9.31 in the stress treatment group and 9.48 in the control group and the difference is not significant (a rank sum test, $p = 0.291$).[20]

For female subjects, we conclude that psycho-social stress lowers willingness to compete because women under stress respond negatively to tournament incentives. Based on this, women react by investing less in the tournament incentive scheme when given a choice. For men, however, we do not find strong evidence of any channel in particular. By process of elimination, we conclude that lower willingness to compete among men is driven by preferences for engaging in competition under stress.

3.5 Conclusion

We experimentally induce stress in the laboratory using a modified TSST-G protocol and find that subjects in the stress treatment group are subsequently less competitive, investing less in the tournament compensation scheme than those in the control group. However, this is only true when the willingness to compete decision is made ex-ante, before the competitive task. In the tasks for which subjects made ex-post willingness to compete decisions, we find no treatment effect. This is true when deciding how much to invest in the tournament compensation scheme for past piece-rate performance both before and under the stress/control procedure. Together, this indicates that stress reduces preferences for performance under competition or subjective beliefs about performance under competition, rather than willingness to compete. We confirm that the treatment

should be interpreted with caution.

[20]The difference is higher for the female sub-sample (9.09 vs. 9.48, $p = 0.163$) than for the male sub-sample (5.53 vs. 9.52, $p = 0.916$). However, we should note that this is only a self-reported comprehension measure, which may reflect other factors, such as confidence.

difference in the ex-ante willingness to compete is caused by a physiological stress reaction by examining salivary cortisol levels.

Perhaps our most important result is that women perform much worse when paid according to the tournament compensation scheme in the stress treatment, compared to women in the control group. The related confidence for tournaments under stress is also lower. Interestingly, we do not observe such a drop in performance or confidence when women are asked to perform under stress, but without competitive incentives. Also, when asked to compete outside stress, women are able to improve their performance, as observed in our control group. It is the combination of stress and tournament incentives which is detrimental to performance. For women, this drop in performance under competition and related confidence can explain the drop in the ex-ante willingness to compete we observe in the stress treatment group. We do not find such a link among men, whose performance and confidence is not affected by the stress treatment. The lower ex-ante willingness to compete among men in the stress treatment seems to be driven by changed preferences for performance under competition, which might be linked to the "flight" reaction to stress.

Our findings could explain past results regarding the effect of tournament incentives on performance, where sometimes a positive effect is found for both genders, but sometimes only for men (Niederle and Vesterlund 2007; Gneezy, Niederle, and Rustichini 2003; Gneezy and Rustichini 2004). Potentially, the competitions differed in the degree of stress involved. Our results also support the claim of Niederle and Vesterlund (2010) that gender gaps in math test-scores may not necessarily reflect differences in math ability. Especially when test results come from highly-competitive and stressful settings, such as university entrance exams, women's performance may fall far below their ability.

Overall, the results presented in this article can help explain the under-representation of women in highly competitive positions. Many competitive situations that affect one's career trajectory—such as exams, job interviews and asking for a promotion—are stressful. If women perform worse in competitive environments under stress, this will directly affect labor market outcomes, and perhaps dissuade women from entering competitive environments in the first place. A question then is which aspects of school or work environments are stressful for women and whether some could be mitigated, possibly by providing psychological support or behavioral training. Also, it is important to discuss whether competitive incentive schemes are not used excessively in our society, especially when targeting women.

Our results show that women are not worse at coping either with stress or with

competition separately, but do significantly worse than men when faced with both. These findings have implications for labor market policy. Tournament incentives and stress are often used to boost performance in firms, and while this incentive structure may be tailor made to suit men, it may be detrimental for both employers when applied to women, as well as for female employees. Moreover, hiring practices might be improved by recognizing that performance in stressful environments may not accurately reflect women's abilities in non-stressful settings.

Increasing the presence of women in a variety of high-calibre careers is widely recognized as an important policy goal. While many fields are inherently both stressful and competitive, one way of closing the gender gap may be to better prepare women for these situations, by focusing on competition under stress. For example, repeated exposure to a stressor has been shown to reduce the magnitude of the stress response (Wüst et al. 2005). Given this, education and training programs targeted at women and girls might concentrate on acclimatizing them better for the types of stress one typically encounters on the labor market, and in developing non-cognitive skills for making work life less stressful.

Figure 3.1: Timeline of the experiment

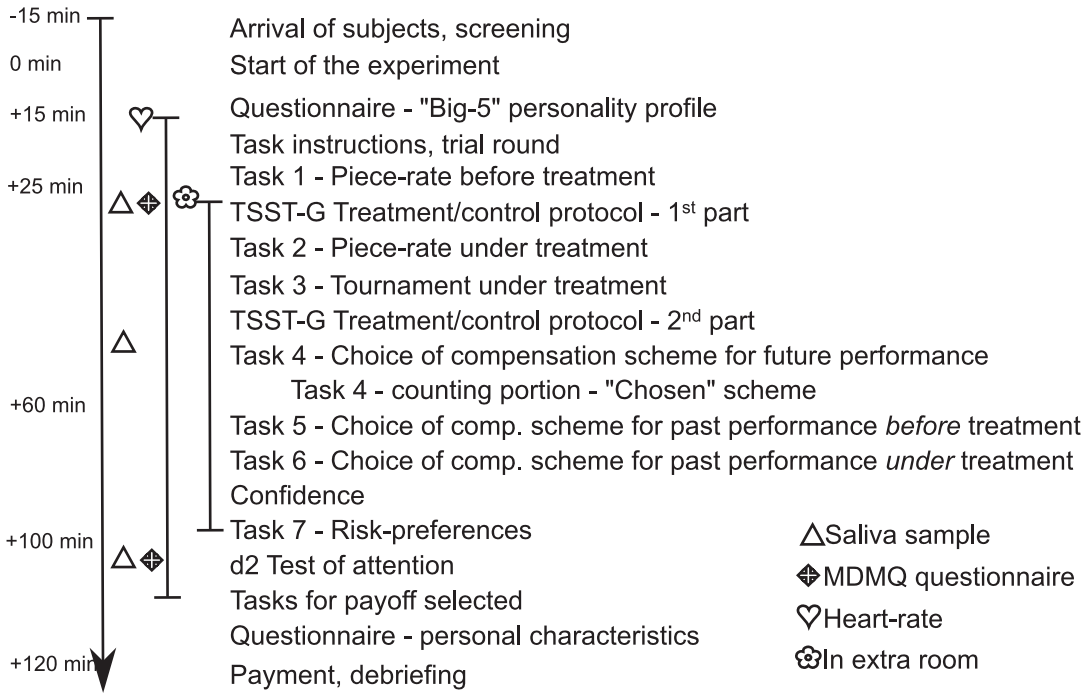
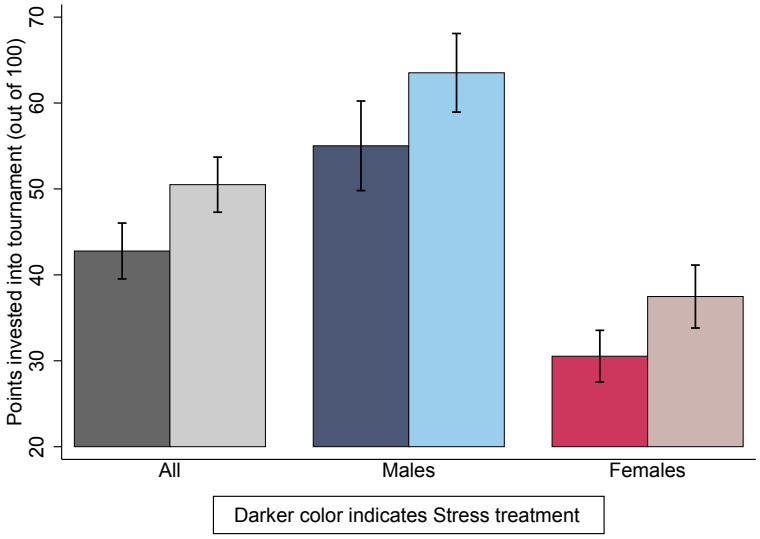
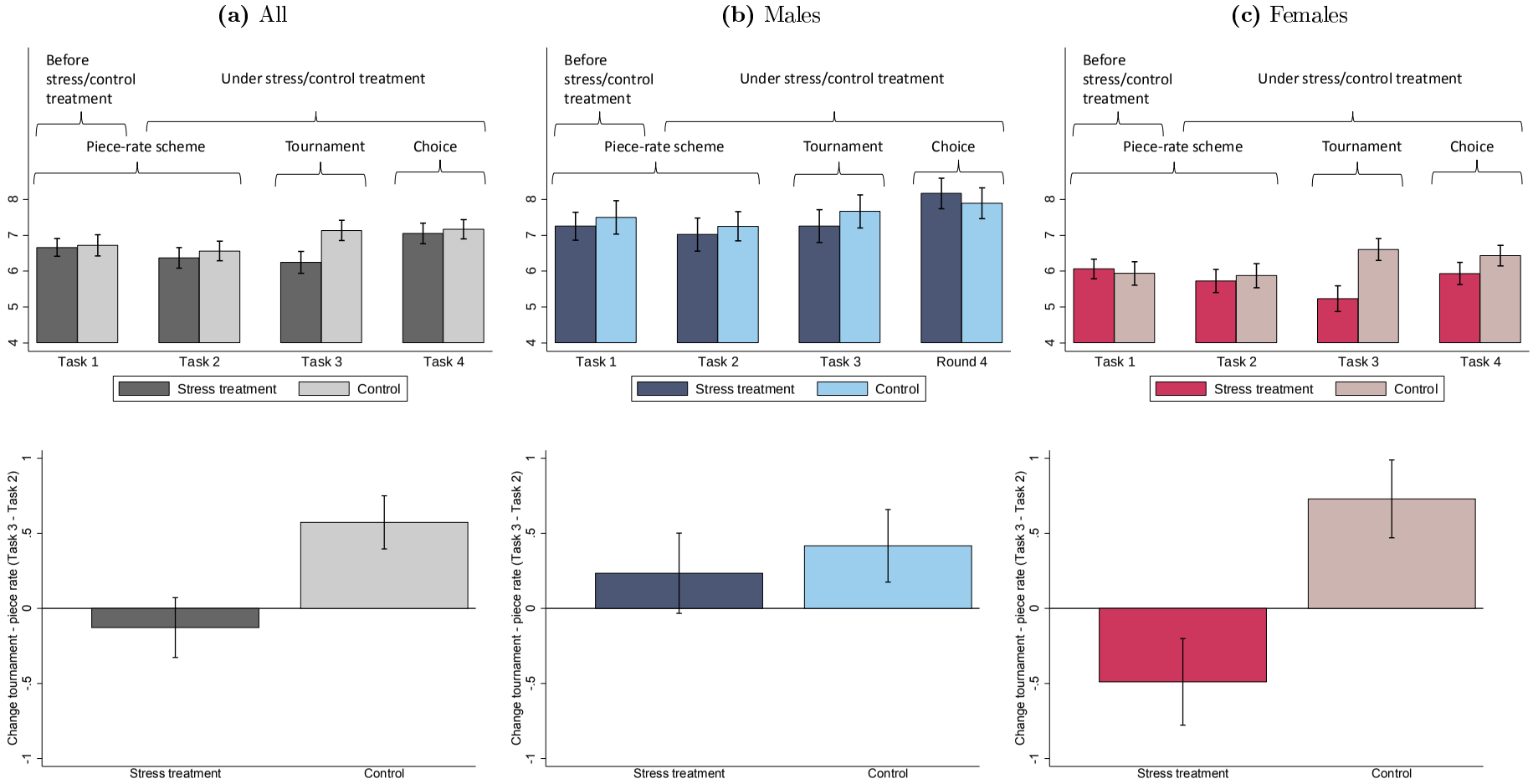


Figure 3.2: Willingness to compete for future performance, by treatment



Notes: Mean willingness to compete for future performance, by treatment. Willingness to compete is measured as the investment in the tournament compensation scheme in Task 4, where the choice occurred before completing the counting portion of the task. 0 indicates all points invested in the piece-rate scheme, and 100 indicates all points invested in the tournament compensation scheme. The darker color indicates the "Stress treatment", i.e. that the subject was exposed to the stressor in the form of a modified TSST-G stress procedure, while the lighter color represents the control group. The bars indicate mean ± standard error.

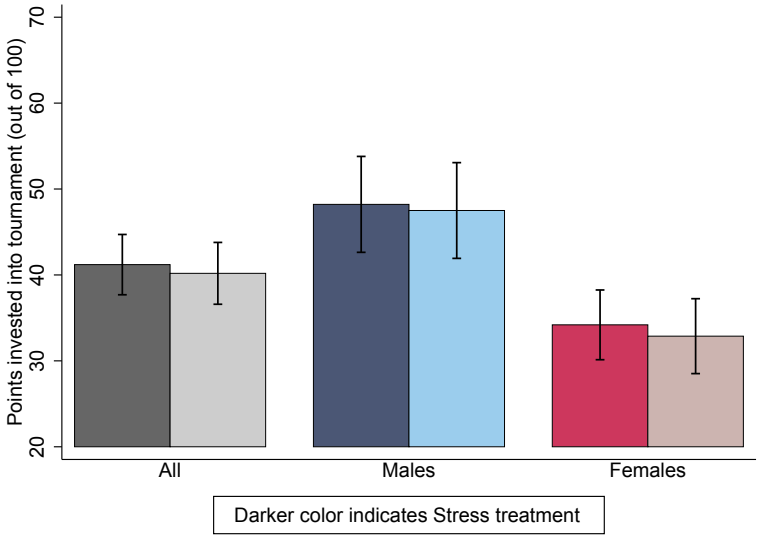
Figure 3.3: Performance under different incentive schemes, by treatment and gender



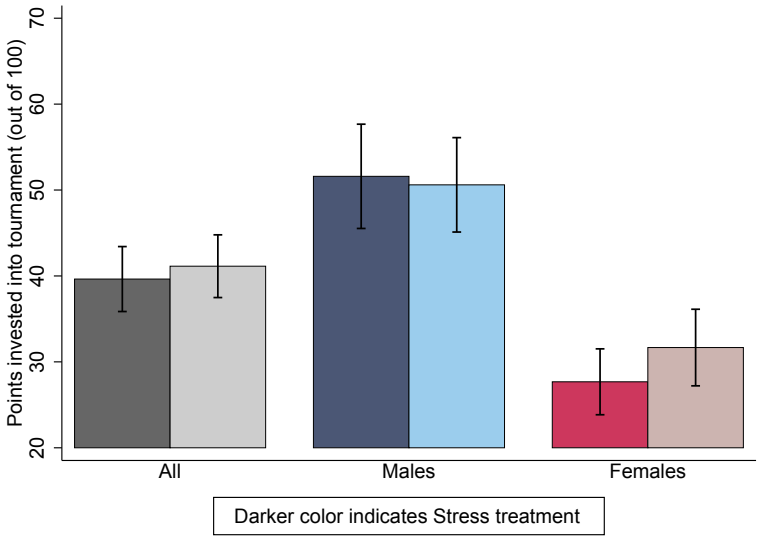
Notes: Performance in the experiment, by gender. "Stress treatment" indicates that the subject was exposed to the stressor in the form of a modified TSST-G stress procedure. The control group was exposed to the TSST-G control procedure. Task 1 is the baseline and was run under piece-rate incentives. Task 2 and Task 3 took place after the first part of the TSST-G stress treatment/control protocol, under piece-rate and tournament incentives, respectively. Task 4 occurred after the second part of the TSST-G stress treatment/control protocol, after the subjects had chosen their preferred incentives scheme for that round, investing 0-100 points into the tournament. Graphs in the bottom panel depict the change between Task 3 and Task 2, i.e. capture the effects of tournament incentives. Bars indicate mean ± standard error.

Figure 3.4: Ex-post willingness to compete, by treatment

(a) Willingness to compete for past performance before treatment

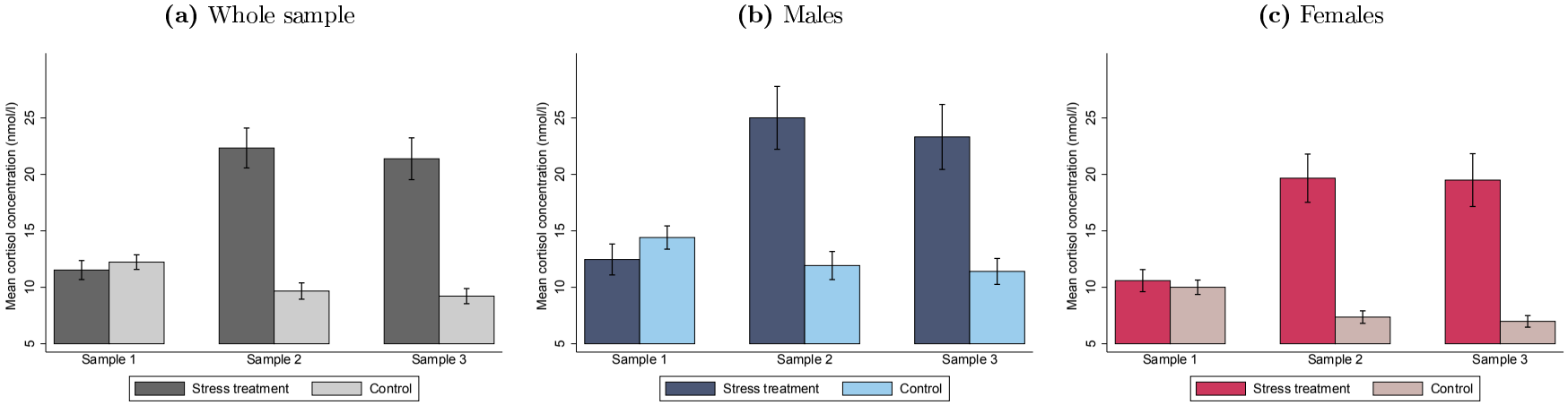


(b) Willingness to compete for past performance under treatment



Notes: Mean willingness to compete for past performance, by stress treatment. Willingness to compete is measured as the ex-post investment in the tournament compensation, regarding the performance that occurred either before the stress/control procedure (Task 5, Panel A), or under the stress/control procedure (Task 6, Panel B), where 0 indicates all points invested in the piece-rate scheme, and 100 indicates all points invested in the tournament compensation scheme. The darker color indicates the "Stress treatment", i.e. that the subject was exposed to the stressor in the form of a modified TSST-G stress procedure, while the lighter color represents the control group. The bars indicate mean $\pm$ standard error.

Figure 3.5: Induced cortisol response, by treatment



Notes: Evolution of the mean salivary cortisol concentration (in nmol/l) over the experiment, by treatment. "Stress treatment" indicates that the subject was exposed to the stressor in the form of a modified TSST-G stress procedure. The control group was exposed to the TSST-G control procedure. Sample 1 was collected prior to the TSST-G stress/control procedure. Sample 2 was collected after the second part of the TSST-G protocol (the counting task) and Sample 3 was collected after Task 7, when subjects returned to the laboratory. For details please consult the timeline in Figure 3.1. The bars indicate mean ± standard error.

Table 3.1: Summary statistics, willingness to compete by treatment and task

	Number of points invested in competition (out of 100)					
		Treatment group			Ranksum	
Sample	*All*	*Stress*	*Control*	Diff.	(p-value)	N
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Task 4 Choice - Future performance under treatment						
All	46.68	42.78	50.50	-7.72	0.046	190
Male	59.32	55.02	63.52	-8.50	0.190	95
Female	34.04	30.53	37.48	-6.95	0.159	95
Panel B: Task 5 Choice - Past performance before treatment						
All	40.69	41.20	40.19	1.01	0.826	190
Male	47.85	48.21	47.50	0.71	0.988	95
Female	33.53	34.19	32.88	1.32	0.688	95
Panel C: Task 6 Choice - Past performance under treatment						
All	40.39	39.64	41.14	-1.50	0.702	190
Male	51.09	51.60	50.60	0.99	0.967	95
Female	29.69	27.68	31.67	-3.99	0.710	95

Notes: Mean decisions regarding willingness to compete, across tasks, treatments, and gender. Panel A presents the ex-ante competitiveness decision, while Panels B and C present the two ex-post competitiveness decisions. "Stress" treatment indicates that the subject was exposed to the stressor in the form of a modified TSST-G procedure. Subjects in the "Control" treatment went through a TSST-G control procedure. All differences are tested using a Wilcoxon rank-sum test.

Table 3.2: Willingness to compete for future performance

Dep. Variable	*Willingness to Compete* *points invested into tournament ex-ante (Task 4)*			
Sample	*All*		*Males*	*Females*
	(1)	(2)	(3)	(4)
Stress treatment	-7.59**	-7.93	-7.99	-7.31
	(3.14)	(5.11)	(5.00)	(4.58)
Female	-22.06***	-22.41***		
	(4.14)	(6.23)		
Stress treatment*Female		0.69		
		(7.38)		
Solved Task 1 (baseline)	2.33**	2.32**	2.07*	2.85**
	(0.95)	(0.95)	(1.15)	(1.34)
Constant	45.89***	46.08***	48.02***	20.55**
	(8.05)	(8.33)	(9.53)	(8.06)
Observations	190	190	95	95
R-squared	0.21	0.21	0.05	0.09

Notes: OLS. Standard errors are clustered on a session level. *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is investment in the tournament compensation scheme in Task 4, where the choice occurred before completing the counting portion of the task. 0 indicates all points invested in the piece-rate scheme, and 100 indicates all points invested in the tournament compensation scheme. "Stress treatment" is a dummy variable indicating that the subject was exposed to the stressor in the form of a modified TSST-G procedure.

Table 3.3: Summary statistics, performance in the counting task

	Number of problems solved correctly					
		Treatment group			Ranksum	
Sample	*All*	*Stress*	*Control*	Diff.	(p-value)	N
	(1)	(2)	(3)	(4)	(5)	
Task 1 - Piece-rate before treatment						
All	6.69	6.66	6.72	-0.06	0.931	190
Male	7.38	7.26	7.50	-0.24	0.943	95
Female	6.00	6.06	5.94	0.13	0.880	95
Task 2 - Piece-rate under treatment						
All	6.47	6.37	6.56	-0.19	0.560	190
Male	7.14	7.02	7.25	-0.23	0.761	95
Female	5.80	5.72	5.88	-0.15	0.632	95
Task 3 - Tournament under treatment						
All	6.69	6.24	7.14	-0.89	0.018	190
Male	7.46	7.26	7.67	-0.41	0.562	95
Female	5.93	5.23	6.60	-1.37	0.003	95
Task 4 - Chosen scheme under treatment						
All	7.11	7.05	7.17	-0.11	0.650	190
Male	8.03	8.17	7.90	0.27	0.538	95
Female	6.19	5.93	6.44	-0.50	0.087	95

Notes: Mean performance in the tasks under different compensation schemes, by treatment and gender. "Stress" treatment indicates that the subject was exposed to the stressor in the form of a modified TSST-G procedure. Subjects in the "Control" treatment went through a TSST-G control procedure.

Table 3.4: The effect of stress, competition and gender on performance

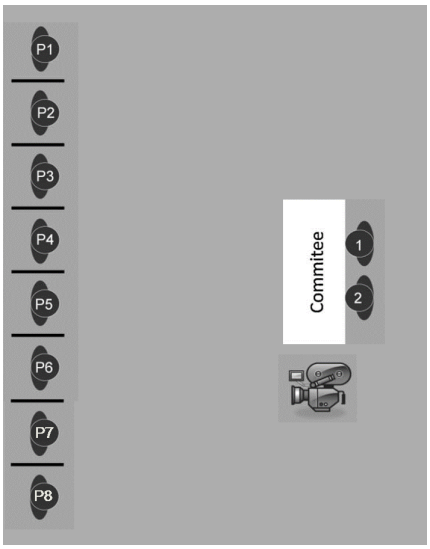
Dep. Variable	*Problems Solved Correctly*							
	Task 3				*Task 3 - Task 2*			
Incentive scheme	*Tournament Incentives*				*Effect of Tournament Incentives*			
Sample	*All*	*All*	*Males*	*Females*	*All*	*All*	*Males*	*Females*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Stress treatment	-0.84***	-0.21	-0.19	-1.45***	-0.70***	-0.17	-0.15	-1.21***
	(0.21)	(0.29)	(0.28)	(0.34)	(0.24)	(0.30)	(0.28)	(0.33)
Female	-0.41	0.22			-0.14	0.39		
	(0.27)	(0.34)			(0.28)	(0.36)		
Stress treatment*Female		-1.26**				-1.05**		
		(0.50)				(0.43)		
Solved Task 1 (baseline)	0.81***	0.82***	0.91***	0.63***	0.04	0.05	0.11**	-0.09
	(0.04)	(0.04)	(0.05)	(0.12)	(0.06)	(0.06)	(0.05)	(0.10)
Constant	1.88***	1.53***	0.85	2.85***	0.34	0.05	-0.44	1.24**
	(0.36)	(0.38)	(0.50)	(0.75)	(0.48)	(0.51)	(0.52)	(0.53)
Observations	190	190	95	95	190	190	95	95
R-squared	0.61	0.62	0.72	0.39	0.04	0.06	0.04	0.10

Notes: OLS. Standard errors are clustered on a session level. *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is the number of addition problems that were correctly completed within the time limit in the specified task. Both Task 2 and Task 3 were completed under treatment, under piece-rate and tournament incentives, respectively. "Stress treatment" is a dummy variable indicating that the subject was exposed to the stressor in the form of a modified TSST-G procedure.

3.A Appendix 3

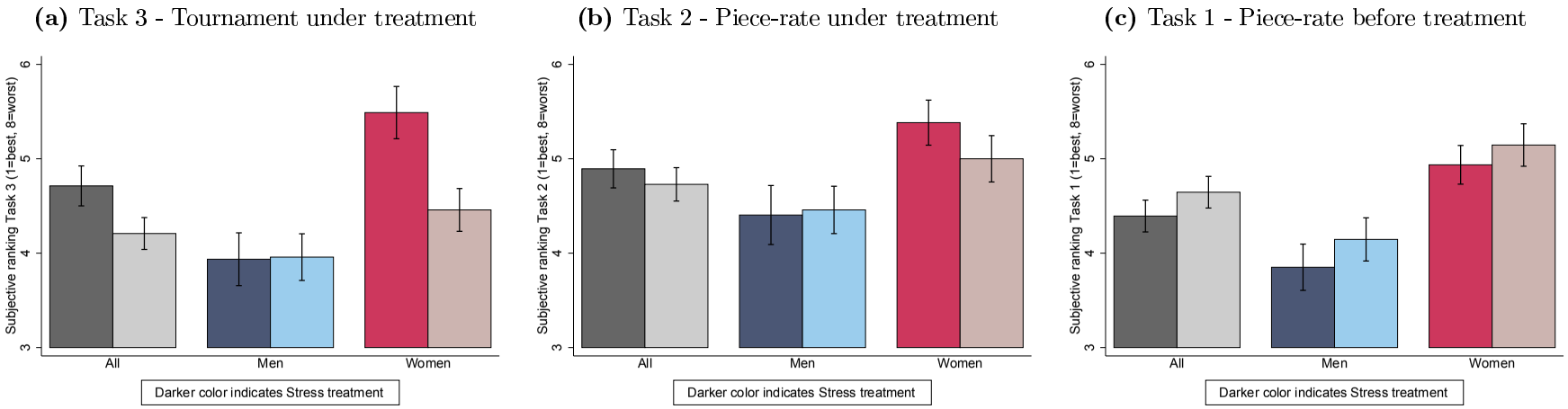
Figure 3.6: TSST-G stress procedure: setting of the room

(a) Schema of the room



(b) Working stations of subjects in the room

Figure 3.7: Confidence: beliefs about the subjects' perceived rank in the session, by treatment and task



Notes: Mean beliefs about the subjects' rank among the 8 subjects in the session, by task and treatment (i.e. 1=highest confidence, 8=lowest confidence). "Stress treatment" indicates that the subject was exposed to the stressor in the form of a modified TSST-G stress procedure. The control group was exposed to the TSST-G control procedure. Task 3 was conducted under tournament incentives under the stress/control treatment, Task 2 was conducted under piece-rate incentives under treatment and Task 1 was completed under piece-rate incentives before the stress/control manipulation. Confidence questions were non-incentivized and were elicited after Task 6. For details please consult the timeline in Figure 3.1. The bars indicate mean ± standard error.

Table 3.5: Willingness to compete using a binary measure

Dep. Variable	*Marginal fixed effects after probit Probability of investing more than 50 points in the tournament*			
Sample	*All*	*All*	*Males*	*Females*
	(1)	*(2)*	*(3)*	*(4)*
Stress treatment	-0.22***	-0.15**	-0.15**	-0.24***
	-0.06	-0.07	-0.07	-0.09
Female	-0.30***	-0.22**		
	-0.07	-0.1		
Stress treatment*Female		-0.17		
		-0.12		
Solved Task 1 (baseline)	0.04***	0.05***	0.04**	0.05***
	-0.02	-0.02	-0.02	-0.02
Observations	190	190	95	95

Notes: Marginal fixed effects after probit. Standard errors are clustered on a session level. *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is a dummy indicating that the subject in Task 4 invested at least 50 points into the tournament compensation scheme, where the choice occurred before completing the counting portion of Task 4. "Stress treatment" is a dummy variable indicating that the subject was exposed to the stressor in the form of a modified TSST-G procedure.

Table 3.6: The effect of stress and gender on performance

Dep. Variable	*Problems Solved Correctly*							
	Task 1		*Task 2*		*Task 2- Task 1*			
Incentive scheme	*Piece-rate*		*Piece-rate*		*Effect of Stress Treatment*			
Sample	*All*	*All*	*All*	*All*	*Al*	*All*	*Males*	*Females*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Stress treatment	-0.06	-0.24	-0.14	-0.04	-0.13	0.02	0.02	-0.28
	(0.39)	(0.75)	(0.26)	(0.33)	(0.28)	(0.40)	(0.40)	(0.36)
Female	-1.38***	-1.56**	-0.28	-0.17	0.04	0.19		
	(0.43)	(0.74)	(0.25)	(0.31)	(0.25)	(0.37)		
Stress treatment*Female		0.37		-0.21		-0.29		
		(0.86)		(0.46)		(0.51)		
Solved Task 1 (baseline)			0.77***	0.77***				
			(0.05)	(0.05)				
Constant	7.41***	7.50***	1.53***	1.48***	-0.18	-0.25	-0.25	-0.06
	(0.51)	(0.67)	(0.40)	(0.43)	(0.26)	(0.31)	(0.31)	(0.22)
Observations	190	190	190	190	190	190	95	95
R-squared	0.07	0.07	0.58	0.58	0.00	0.00	0.00	0.01

Notes: OLS. Standard errors are clustered on a session level. *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is the number of addition problems that were correctly completed within the time limit in the specified task. Both Task 1 and Task 2 were completed under piece-rate incentives, before and after the stress/control treatment, respectively. "Stress treatment" is a dummy variable indicating that the subject was exposed to the stressor in the form of a modified TSST-G procedure.

Table 3.7: Performance under the chosen compensation scheme, across treatments

Dep. Variable	*Problems Solved Correctly Task 4*			
Incentive scheme	*Chosen Incentives Scheme*			
Sample	*All*	*All*	*Males*	*Females*
	(1)	(2)	(3)	(4)
Stress Treatment	-0.84***	-0.21	-0.19	-1.45***
	(0.21)	(0.29)	(0.28)	(0.34)
Female	-0.41	0.22		
	(0.27)	(0.34)		
Stress Treatment*Female		-1.26**		
		(0.50)		
Solved Task 1 (baseline)	0.81***	0.82***	0.91***	0.63***
	(0.04)	(0.04)	(0.05)	(0.12)
Constant	1.88***	1.53***	0.85	2.85***
	(0.36)	(0.38)	(0.50)	(0.75)
Observations	190	190	95	95
R-squared	0.61	0.62	0.72	0.39

Notes: OLS. Standard errors are clustered on a session level. *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is the number of addition problems that were correctly completed within the time limit in the counting portion of Task 4, before which the subjects were asked to chose their preferred compensation scheme, choosing any linear combination of a piece-rate compensation scheme and a tournament compensation scheme. "Stress Treatment" is a dummy variable indicating that the subject was exposed to the stressor in the form of a modified TSST-G procedure.

Table 3.8: Willingness to compete for past performance before and under treatment

Dep. Variable	*Willingness to Compete*							
	points invested into tournament ex-post							
	for Task 1 (piece-rate before Treatment)				*for Task 2 (piece-rate under Treatment)*			
Sample	*All*	*All*	*Males*	*Females*	*All*	*All*	*Males*	*Females*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Stress Treatment	1.30	1.89	2.24	1.07	-1.34	1.63	1.87	-4.07
	(4.81)	(6.01)	(6.32)	(5.67)	(4.37)	(6.59)	(6.88)	(5.28)
Female	-7.68*	-7.09			-17.82***	-14.85**		
	(4.02)	(7.16)			(4.40)	(6.01)		
Stress Treatment*Female		-1.19				-5.95		
		(7.07)				(7.94)		
Solved Task 1 (baseline)	4.82***	4.82***	6.25***	1.93	2.60**	2.62**	3.58**	0.67
	(1.03)	(1.03)	(0.98)	(1.70)	(1.01)	(1.02)	(1.34)	(1.18)
Constant	11.64	11.32	0.64	21.43*	32.60***	30.97***	23.78**	27.71***
	(10.00)	(10.90)	(9.52)	(11.28)	(8.03)	(8.65)	(10.63)	(7.03)
Observations	190	190	95	95	190	190	95	95
R-squared	0.17	0.17	0.23	0.02	0.12	0.12	0.07	0.01

Notes: OLS. Standard errors are clustered on a session level. *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is the ex-post investment in the tournament compensation, regarding the performance that occurred either before the stress/control procedure (Task 5, Columns 1-4), or under the stress/control procedure (Task 6, Columns 5-8), where 0 indicates all points invested in the piece-rate scheme, and 100 indicates all points invested in the tournament compensation scheme. "Stress Treatment" indicates that the subject was exposed to the stressor in the form of a modified TSST-G procedure.

Table 3.9: Willingness to compete for future performance, by cortisol response

Panel A: OLS and IV (2SLS) Estimation

Dep. Variable	Willingness to Compete points invested into tournament ex ante (Task 4)					
Sample	*All*		*Males*		*Females*	
Estimation Method	*OLS*	*IV*	*OLS*	*IV*	*OLS*	*IV*
	(1)	(2)	(3)	(4)	(5)	(6)
Cortisol response	-2.35	-5.62**	-0.90	-5.64	-3.48***	-5.65
	(1.52)	(2.31)	(2.91)	(3.66)	(0.88)	(3.58)
Female	-22.39***	-22.92***				
	(4.28)	(4.29)				
Solved Task 1 (baseline)	2.33**	2.32**	2.13*	2.17*	2.68**	2.61**
	(0.95)	(0.94)	(1.13)	(1.15)	(1.28)	(1.25)
Constant	43.49***	45.54***	44.12***	46.62***	19.49**	20.85**
	(8.33)	(7.99)	(9.96)	(9.26)	(7.68)	(7.70)
Observations	189	189	95	95	94	94
R-squared	0.20	0.19	0.04	0.00	0.11	0.09

Panel B: First-stage regressions

Dep. Variable	Cortisol response: (sample 2-1)/sample 1		
Sample	*All*	*Males*	*Females*
Estimation Method	*OLS*	*OLS*	*OLS*
	(1)	(2)	(3)
Stress Treatment	1.36***	1.42***	1.32***
	(0.14)	(0.18)	(0.26)
Female	-0.17		
	(0.15)		
Solved Task 1 (baseline)	-0.00	0.02	-0.04
	(0.03)	(0.04)	(0.04)
Constant	-0.07	-0.25	0.03
	(0.21)	(0.30)	(0.27)
Observations	189	95	94
R-squared	0.27	0.32	0.22
Cragg-Donald Wald F stat.	66.44	43.6	26.04

Notes: OLS and 2SLS, as indicated by column. Standard errors are clustered on a session level. *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is investment in the tournament compensation scheme in Task 4, where the choice occurred before completing the counting portion of the task. 0 indicates all points invested in the piece-rate scheme, and 100 indicates all points invested in the tournament compensation scheme. "Cortisol response" is measured as a relative increase in the salivary cortisol levels between sample 1 and sample 2 (i.e. (sample 2-sample 1)/sample 1). Sample 1 was collected prior to the TSST-G stress/control procedure, while Sample 2 was collected after the second part of the TSST-G protocol (the counting task). For details regarding the timeline please consult Figure 3.1. "Stress treatment" is a dummy variable indicating that the subject was exposed to the stressor in the form of a modified TSST-G procedure.

Table 3.10: Sensitivity analysis: Willingness to compete and tournament performance

Dependent variable	Willingness to Compete								
	points invested into tournament ex-ante (Task 4)								
Sample	*All*			*Males*			*Females*		
	(1)	(2)	(4)	(4)	(5)	(6)	(7)	(8)	(9)
Solved Task 3		4.38***	4.62***		5.22**	5.34**		4.46***	4.53***
		(0.91)	(0.91)		(1.87)	(1.92)		(0.97)	(0.84)
Stress treatment	-7.59**	-3.89		-7.99	-7.01		-7.31	-0.85	
	(3.14)	(3.13)		(5.00)	(5.13)		(4.58)	(4.72)	
Female	-22.06***	-20.25***	-20.13***						
	(4.14)	(4.36)	(4.41)						
Solved Task 1 (baseline)	2.33**	-1.24	-1.43	2.07*	-2.68	-2.74	2.85**	0.03	-0.02
	(0.95)	(1.15)	(1.15)	(1.15)	(2.15)	(2.17)	(1.34)	(1.59)	(1.53)
Constant	45.89***	37.67***	35.33***	48.02***	43.60***	39.68***	20.55**	7.87	7.31
	(8.05)	(8.00)	(7.93)	(9.53)	(9.35)	(9.19)	(8.06)	(9.36)	(8.65)
Observations	190	190	190	95	95	95	95	95	95
R-squared	0.21	0.27	0.27	0.05	0.11	0.10	0.09	0.21	0.21
Adjusted R-squared	0.20	0.26	0.26	0.03	0.09	0.08	0.07	0.18	0.19

Notes: OLS. Standard errors are clustered on a session level. *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is investment in the tournament compensation scheme in Task 4, where the choice occurred before completing the counting portion of the task. 0 indicates all points invested in the piece-rate scheme, and 100 indicates all points invested in the tournament compensation scheme. "Solved Task 3" captures the number of correctly answered problems in Task 3, which occurred under tournament incentives under treatment. "Stress treatment" is a dummy variable indicating that the subject was exposed to the stressor in the form of a modified TSST-G procedure.

Table 3.11: Sensitivity analysis: Tournament performance and related confidence

Dependent variable	Confidence in Task 3 Perceived rank among 8 subjects in session								
Sample	*All*			*Males*			*Females*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Stress Treatment	0.49**	0.03		-0.08	-0.18		1.05***	0.26	
	(0.20)	(0.17)		(0.31)	(0.28)		(0.29)	(0.29)	
Female	0.72***	0.49**	0.49**						
	(0.25)	(0.23)	(0.23)						
Solved Task 3		-0.55***	-0.55***		-0.54***	-0.53***		-0.55***	-0.57***
		(0.05)	(0.04)		(0.08)	(0.08)		(0.05)	(0.04)
Solved Task 1 (baseline)	0.22***	0.23***	0.23***	-0.25***	0.24**	0.23**	-0.16**	0.18**	0.20**
	(0.04)	(0.05)	(0.05)	(0.04)	(0.09)	(0.09)	(0.07)	(0.08)	(0.08)
Constant	5.31***	6.33***	6.35***	5.84***	6.30***	6.19***	5.42***	6.98***	7.15***
	(0.34)	(0.34)	(0.32)	(0.40)	(0.38)	(0.36)	(0.46)	(0.54)	(0.48)
Observations	190	190	190	95	95	95	95	95	95
R-squared	0.18	0.46	0.46	0.17	0.41	0.41	0.12	0.42	0.42
Adjusted R-squared	0.17	0.45	0.45	0.15	0.39	0.40	0.10	0.41	0.41

Notes: OLS. Standard errors are clustered on a session level. *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is confidence in Task 3, measured as subjects' subjective ranking among 8 participants in the session. 1 thus indicates highest confidence and 8 the lowest confidence. Task 3 occurred under tournament incentives under treatment."Stress Treatment" indicates that the subject was exposed to the stressor in the form of a modified TSST-G procedure.

Bibliography

Abbink, Klaus, and Benedikt Herrmann. 2011. "The Moral Costs of Nastiness." *Economic Inquiry* 49 (2): 631–633.

Abbink, Klaus, and Abdolkarim Sadrieh. 2009. "The pleasure of being nasty." *Economics Letters* 105 (3): 306–308.

Akerlof, George A., and Rachel E. Kranton. 2000. "Economics and Identity." *Quarterly Journal of Economics* 115 (3): 715–753.

Alesina, Alberto, and Eliana La Ferrara. 2002. "Who trusts others?" *Journal of Public Economics* 85 (2): 207–234.

Allport, G W. 1954. *The nature of prejudice*. Reading, MA: Addison Wesley.

Almas, Ingvild, Alexander W. Cappelen, Erik O. Sorensen, and Bertil Tungodden. 2010. "Fairness and the development of inequality acceptance." *Science* 328 (5982): 1176–8.

Angerer, Silvia, Daniela Glätzle-Rützler, Philipp Lergetporer, and Matthias Sutter. 2015. "Cooperation and Discrimination Within and Across Language Borders: Evidence from Children in a Bilingual City." *IZA Disscussion Papers*, no. 9039.

Arrow, Kenneth J. 1973. "The Theory of Discrimination." In *Discrimination in Labor Markets*, edited by Orley Ashenfelter and Albert Rees, 3–33. Princeton University Press.

———. 1998. "What Has Economics to Say About Racial Discrimination ?" *Journal of Economic Perspectives* 12 (2): 91–100.

Banerjee, Abhijit V. 1992. "A Simple Model of Herd Behavior." *Quarterly Journal of Economics* 107 (3): 797–817.

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

Bardsley, Nicholas. 2005. "Experimental economics and the artificiality of alteration." *Journal of Economic Methodology* 12 (2): 239–251.

Bartoš, Vojtěch, Michal Bauer, Julie Chytilová, and Filip Matějka. 2015. "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition." *American Economic Review*, no. forthcoming.

Basu, Kaushik. 2005. "Racial conflict and the malignancy of identity." *The Journal of Economic Inequality* 3 (3): 221–241.

Bauer, Michal, Alessandra Cassar, Julie Chytilová, and Joseph Henrich. 2014. "War's enduring effects on the development of egalitarian motivations and in-group biases." *Psychological Science* 25 (1): 47–57.

Bauer, Michal, Nathan Fiala, and Ian V. Levely. 2014. "Trusting Former Rebels: An Experimental Approach to Understanding Reintegration after Civil War." *CERGE-EI Working Paper Series*, vol. No. 512.

Baum, Andrew, and Neil E. Grunberg. 1997. "Measuring Stress Hormones." Chapter Chapter 8 of *Measuring Stress: A Guide for Health and Social Scientists*, edited by Sheldon Cohen, Ronald C. Kessler, and Lynn Underwood Gordon, 175–192. Oxford: Oxford University Press.

Baumeister, Roy F. 1995. *Evil: Inside Human Cruelty and Violence*. New York: W.H. Freeman.

Bayer, Patrick, Randi Hjalmarsson, and David Pozen. 2009. "Building Criminal Capital Behind Bars: Peer Effects in Juvenile Corrections." *Quarterly Journal of Economics* 124 (1): 105–147.

Beaman, Lori, Raghabendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova. 2009. "Powerful women: does exposure reduce bias?" *Quarterly Journal of Economics* 124 (4): 1497–1540.

Becker, Gary S. 1971. *The Economics of Discrimination*. University of Chicago Press.

Bellows, John, and Edward Miguel. 2009. "War and Local Collective Action in Sierra Leone." *Journal of Public Economics* 93 (11): 1144–1157.

Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. "Trust, Reciprocity, and Social History." *Games and Economic Behavior* 10 (1): 122–142.

Berge, Ivar Lars Oppedal, Kjetil Bjorvatn, Simon Galle, Edward Miguel, Daniel Posner, Bertil Tungodden, and Kelly Zhang. 2015a. "How strong are ethnic preferences." *Mimeo*.

Berge, Lars Ivar Oppedal, Kjetil Bjorvatn, Armando Jose Garcia Pires, and Bertil Tungodden. 2015b. "Competitive in the lab, successful in the field?" *Journal of Economic Behavior & Organization* 118:303–317.

Bernhard, Helen, Ernst Fehr, and Urs Fischbacher. 2006. "Group Affiliation and Altruistic Norm Enforcement." *American Economic Review* 96 (2): 217–221.

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

Bertrand, Marianne. 2009. "CEOs." *Annual Review of Economics* 1 (1): 121–150.

Bertrand, Marianne, and Kevin F. Hallock. 2001. "The gender gap in top corporate jobs." *Industrial and Labor Relations Review* 55 (1): 3–21.

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.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. 1992. "A Theory of Fads , Fashion , Custom , and Cultural Change as Informational Cascades." *Journal of Political Economy* 100 (5): 992–1026.

———. 1998. "Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades." *Journal of Economic Perspectives* 12 (3): 151–170.

Blattman, Christopher. 2009. "From Violence to Voting: War and political participation in Uganda." *American Political Science Review* 103 (02): 231–247.

Boisjoly, Johanne, Greg J Duncan, Michael Kremer, Dan M Levy, and Jacque Eccles. 2006. "Empathy or Antipathy? The Impact of Diversity." *American Economic Review* 96 (5): 1890–1905.

Bonnier, Evelina, Jonas Poulsen, Thorsten Rogall, and Miri Stryjan. 2015. "Preparing for Genocide: Community Work in Rwanda." *Mimeo*.

Bornhorst, Fabian, Andrea Ichino, Oliver Kirchkamp, Karl H. Schlag, and Eyal Winter. 2010. "Similarities and differences when building trust: the role of cultures." *Experimental Economics* 13 (3): 260–283.

Bowles, Samuel. 2008. "Conflict: Altruism's Midwife." *Nature* 456 (20): 326–327.

Brickenkamp, Rolf, and Eric Zillmer. 1998. *The d2 test of attention*. Hogrefe & Huber Pub.

Buchan, Nancy R., Eric J. Johnson, and Rachel T. A. Croson. 2006. "Let's get personal: An international examination of the influence of communication, culture and social distance on other regarding preferences." *Journal of Economic Behavior & Organization* 60 (3): 373–398.

Buckert, Magdalena, Christiane Schwieren, Brigitte M Kudielka, and Christian J Fiebach. 2015. "How stressful are economic competitions in the lab? An investigation with physiological measures." *Discussion Paper Series, University of Heidelberg, Department of Economics*, no. No. 0593.

Buser, Thomas, Anna Dreber, and Johanna Mollerstrom. 2015. "Stress Reactions cannot explain the Gender Gap in Willingness to compete." *Tinbergen Institute Discussion Paper*, no. 15-059/I.

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek. 2014. "Gender, Competitiveness and Career Choices." *Quarterly Journal of Economics* 129 (3): 1409–1447.

Cahlíková, Jana. 2015. "Study Abroad Experience and Attitudes Towards Other Nationalities." *CERGE-EI Working Paper Series*, no. 556.

Cahlíková, Jana, and Lubomír Cingl. 2016. "Risk Preferences under Acute Stress." *Experimental Economics*, no. forthcoming.

Cannon, Walter B. 1932. "The wisdom of the body." *American Journal of the Medical Sciences* 184 (6): 864.

Card, David, and Laura Giuliano. 2013. "Peer Effects and Multiple Equilibria in the Risky Behavior of Friends." *Review of Economics and Statistics* 95 (4): 1130–1149.

Charness, Gary, Luca Rigotti, and Aldo Rustichini. 2007. "Individual Behavior and Group Membership." *American Economic Review* 97 (4): 1340–1352.

Chen, Yan, and Sherry Xin Li. 2009. "Group identity and social preferences." *American Economic Review* 99 (1): 431–457.

Choi, Jung-Kyoo, and Samuel Bowles. 2007. "The Coevolution of Parochial Altruism and War." *Science* 318 (5850): 636–640.

Clingingsmith, David, Asim Ijaz Khwaja, and Michael Kremer. 2009. "Estimating the Impact of the Hajj: Religion and Tolerance in Islam's Global Gathering." *Quarterly Journal of Economics* 124 (3): 1133–1170.

Clow, Angela. 2001. "The physiology of stress." In *Stress, myth, theory, and research*, edited by Fiona Jones and Jim Bright, 47–61. Harlow: Pearson Education.

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

Damm, Anna Piil, and Christian Dustmann. 2014. "Does growing up in a high crime neighborhood affect youth criminal behavior?" *American Economic Review* 104 (6): 1806–1832.

DellaVigna, Stefano, Ruben Enikolopov, Vera Mironova, and Maria Petrova. 2014. "Cross-border media and nationalism: Evidence from Serbian radio in Croatia." *American Economic Journal: Applied Economics* 6 (3): 103–32.

Dickerson, Sally S, and Margaret E Kemeny. 2004. "Acute Stressors and Cortisol Responses: A Theoretical Integration and Synthesis of Laboratory Research." *Psychological Bulletin* 130 (3): 355–391.

Di Pietro, Giorgio. 2015. "Do study abroad programs enhance the employability of graduates?" *Education Finance and Policy*.

Dobbie, Will, and Roland G Fryer Jr. 2015. "The Impact of Voluntary Youth Service on Future Outcomes: Evidence from Teach For America." *The BE Journal of Economic Analysis & Policy* 15 (3): 1031–1065.

Dohmen, By Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2010. "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review* 100 (June): 1238–1260.

Dozier, Rush. 2002. *Why We Hate*. Chicago: Contemporary books.

Eisenberg, Daniel, Ezra Golberstein, and Janis L. Whitlock. 2014. "Peer effects on risky behaviors: New evidence from college roommate assignments." *Journal of Health Economics* 33 (1): 126–138.

Esteban, Joan-Maria, and Debraj Ray. 2008. "On the Salience of Ethnic Conflict." *American Economic Review*, no. 1993:2185–2202.

European Commission. 2004. *The situation of Roma in an enlarged European Union*. Luxembourg: Office for Official Publications of the European Communitie.

———. 2014. *The Erasmus Impact Study: Effecs of mobility on the skills and employability of students and the internationalisation of higher education institutions*. European Union.

Falk, Armin, and Christian Zehnder. 2013. "A City-Wide Experiment on Trust Discrimination." *Journal of Public Economics* 100:15–27.

Fearon, James D., and David D Laitin. 2000. "Violence and the Social Construction of Ethnic Identity." *International Organization* 54 (4): 845–877.

Fehr, Ernst, Daniela Glätzle-Rützler, and Matthias Sutter. 2013. "The development of egalitarianism, altruism, spite and parochialism in childhood and adolescence." *European Economic Review* 64:369–383.

Fershtman, Chaim, and Uri 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): 171–178.

Fletcher, Thomas D, Debra Major, and Donald D Davis. 2008. "The interactive relationship of competitive climate and trait competitiveness with workplace attitudes, stress , and performance." *Journal of Organizational Behavior* 922 (August 2006): 899–922.

Flory, Jeffrey A, Andreas Leibbrandt, and John A List. 2015. "Do Competitive Work Places Deter Female Workers? A Large-scale Natural Field Experiment on Gender Differences in Job-entry Decisions." *Review of Economic Studies* 82 (1): 122–155.

FRA, and UNDP. 2012. *The Situation of Roma in 11 EU Member States - Survey Results at a Glance*. Luxem: Publications Office of the European Union.

FXB Center for Health and Human Rights at Harvard University. 2014. "Accelerating Patterns of Anti-Roma Violence in Hungary."

Gächter, Simon, and Benedikt Herrmann. 2011. "The limits of self-governance when cooperators get punished: Experimental evidence from urban and rural Russia." *European Economic Review* 55 (2): 193–210.

Gächter, Simon, Daniele Nosenzo, and Martin Sefton. 2013. "Peer Effects in Pro-Social Behavior: Social Norms or Social Preferences?" *Journal of the European Economic Association* 11 (3): 548–573.

Gentzkow, Matthew A, and Jesse M Shapiro. 2004. "Media , Education and Anti-Americanism in the Muslim World." *Journal of Economic Perspectives* 18 (3): 117–133.

Gilligan, Michael J, Benjamin J Pasquale, and Cyrus D Samii. 2014. "Civil War and Social Cohesion: Lab-in-the-Field Evidence from Nepal." *American Journal of Political Science* 58 (3): 604–619.

Gino, Francesca, Shahar Ayal, and Dan Ariely. 2009. "Contagion and differentiation in unethical behavior: the effect of one bad apple on the barrel." *Psychological science* 20 (3): 393–8.

Glaeser, Edward L. 2005. "The Political Economy of Hatred." *Quarterly Journal of Economics* 120 (1): 45–86.

Glaeser, Edward L., David I. Laibson, Jose A. Scheinkman, and Christine L. Soutter. 2000. "Measuring Trust." *Quarterly Journal of Economics* 115 (3): 811–846.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. 2003. "Performance in Competitive Environments: Gender Differences." *Quarterly Journal of Economics* 118 (3): 1049–1074.

Gneezy, Uri, and Aniela Pietrasz. 2013. "When Half of the Men Are More Competitive than All Women." *Mimeo.*

Gneezy, Uri, and Aldo Rustichini. 2004. "Gender and Competition at a Young Age." *American Economic Review* 94 (2): 377–381.

Goette, Lorenz, Samuel Bendahan, John Thoresen, Fiona Hollis, and Carmen Sandi. 2015. "Stress pulls us apart : Trait anxiety modulates self-confidence under stress." *Psychoneuroendocrinology* 54:115–123.

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): 212–216.

Golkar, Armita, Marie Björnstjerna, and Andreas Olsson. 2015. "Learned fear to social out-group members are determined by ethnicity and prior exposure." *Frontiers in psychology* 6 (February): 123.

Gorodnichenko, Yuriy, and Gerard Roland. 2011. "Which dimensions of culture matter for long-run growth?" *American Economic Review* 101 (3): 492–498.

Greiner, Ben. 2004. "An online recruitment system for economic experiments." In *Forschung und wissenschaftliches Rechnen GWDG Bericht 63*, edited by Kurt Kremer and Volker Macho, 79–93. Gottingen: Ges. fur Wiss. Datenverarbeitung.

Guiso, Luigi, Paola Sapienya, and Luigi Zingales. 2009. "Cultural Biases in Economic Exchange." *Quarterly Journal of Economics* 124 (3): 1095–1131.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Erst Fehr, Herbert Gintis, and Richard Mcelreath. 2001. "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies." *American Economic Review* 91 (2): 73–78.

Henrich, Joseph, Richard Mcelreath, Abigail Barr, Jean Ensminger, Clark Barrett, Alexander Bolyanatz, Juan Camilo Cardenas, Michael Gurven, Edwins Gwako, Natalie Henrich, Carolyn Lesorogol, and Frank an Marlowe. 2006. "Costly Punishment Across Human Societies." *Science* 312 (June): 1767–1770.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter. 2008. "Antisocial punishment across societies." *Science* 319 (5868): 1362–1367.

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

Holm, Hakan J., and Anders Danielson. 2005. "Tropic Trust versus Nordic Trust: Experimental Evidence from Tanzania and Sweden." *Economic Journal* 115 (2005): 505–532.

Jurajda, Stepan, and Daniel Munich. 2011. "Gender Gap in Performance under Competitive Pressure: Admissions to Czech Universities." *American Economic Review: Papers & Proceedings* 101 (3): 514–518.

Kajantie, Eero, and David I W Phillips. 2006. "The effects of sex and hormonal status on the physiological response to acute psychosocial stress." *Psychoneuroendocrinology* 31 (2): 151–178.

Keizer, Kees, Siegwart Lindenberg, and Linda Steg. 2008. "The Spreading of Disorder." *Science* 322 (5908): 1681–1685.

Kirschbaum, Clemens, K-M Pirke, and Dirk H Hellhammer. 1993. "The Trier Social Stress Test - A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting." *Neuropsychobiology* 28 (1-2): 76–81.

Knack, Stephen, and Philip Keefer. 1997. "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation." *Quarterly Journal of Economics* 112 (4): 1251–1288.

Kremer, Michael, and Dan Levy. 2008. "Peer Effects and Alcohol Use among College Students." *Journal of Economic Perspectives* 22 (3): 189–206.

Kudielka, Brigitte M, and Clemens Kirschbaum. 2005. "Sex differences in HPA axis responses to stress: a review." *Biological psychology* 69 (1): 113–132 (May).

Laar, Colette Van, Shana Levin, Stacey Sinclair, and Jim Sidanius. 2005. "The effect of university roommate contact on ethnic attitudes and behavior." *Journal of Experimental Social Psychology* 41 (4): 329–345.

Levin, Jack, and Jack McDevitt. 2002. *Hate Crimes Revisited: America's War Against Those who are Different*. Basic Books.

Mallan, Kimberley M., James Sax, and Ottmar V. Lipp. 2009. "Verbal instruction abolishes fear conditioned to racial out-group faces." *Journal of Experimental Social Psychology* 45 (6): 1303–1307.

McEwen, Bruce S. 2012. "Brain on stress: how the social environment gets under the skin." *Proceedings of the National Academy of Sciences of the United States of America* 109 (Supplement 2): 17180–5.

Mušinka, Alexander, Daniel Škobla, Jakob Hurrle, Kvetoslava Matlovičová, and Jaroslav Kling. 2014. *ATLAS rómskych komunít na Slovensku 2013*. Bratislava: UNDP Bratislava Regional Centre.

Navarrete, Carlos David, Andreas Olsson, Arnold K Ho, Wendy Berry Mendes, Lotte Thomsen, and James Sidanius. 2009. "Fear Extinction to an Out-Group Face The Role of Target Gender." *Psychological Science* 20 (2): 155–158.

Niederle, Muriel, and Lise Vesterlund. 2007. "Do women shy away from competition? Do men compete too much?" *Quarterly Journal of Economics* 122 (August): 1067–1101.

———. 2010. "Explaining the Gender Gap in Math Test Scores: The Role of Competition." *Journal of Economic Perspectives* 24 (2): 129–144.

———. 2011. "Gender and Competition." *Annual Review of Economics* 3 (1): 601–630.

Olsson, Andreas, Jeffrey P Ebert, Mahzarin R Banaji, and Elizabeth a Phelps. 2005. "The role of social groups in the persistence of learned fear." *Science* 309 (5735): 785–7.

Parey, Matthias, and Fabian Waldinger. 2010. "Studying Abroad and the Effect on International Labour Market Mobility: Evidence from the Introduction of Erasmus." *The Economic Journal* 121 (March): 194–222.

Pettigrew, Thomas F. 1998. "Intergroup contact theory." *Annual review of psychology* 49:65–85.

Pettigrew, Thomas F, and Linda R Tropp. 2006. "A meta-analytic test of intergroup contact theory." *Journal of personality and social psychology* 90 (5): 751–83.

Prediger, Sebastian, Bjorn Vollan, and Benedikt Herrmann. 2014. "Resource scarcity and antisocial behavior." *Journal of Public Economics* 119:1–9.

Rammstedt, Beatrice, and Oliver P. John. 2007. "Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German." *Journal of Research in Personality* 41 (1): 203–212.

Rogall, Thorsten. 2015. "Mobilizing the Masses for Genocide." *Mimeo.*

Sacerdote, Bruce I. 2001. "Peer effects with random assignment: results for dartmouth roommates." *Quarterly Journal of Economics* 116 (2): 681–704.

Shayo, Moses, and Asaf Zussman. 2011. "Judicial ingroup bias in the shadow of terrorism." *Quarterly Journal of Economics* 126 (3): 1447–1484.

Spielberger, Charles D., Richard L. Gorsuch, Robert E. Lushene, Peter R. Vagg, and Gerard A. Jacobs. 1983. *Manual for the State-Trait Anxiety Inventory (Form Y) (Self-Evaluation Questionnaire).* Palo Alto: Consulting Psychologists Press.

Starcke, Katrin, and Matthias Brand. 2012. "Decision making under stress: a selective review." *Neuroscience and biobehavioral reviews* 36 (4): 1228–48.

Steyer, Rolf, Peter Schwenkmezger, Peter Notz, and Michael Eid. 1997. "MDBF-Mehrdimensionaler Befindlichkeitsfragebogen." *Gottingen, Germany: Hogrefe.*

Stroud, Laura R, Peter Salovey, and Elissa S Epel. 2002. "Sex differences in stress responses: social rejection versus achievement stress." *Biological Psychiatry* 52 (4): 318–327.

Tabellini, Guido. 2010. "Culture and Institutions: Economic Development in the Regions of Europe." *Journal of the European Economic Association* 8 (June): 677–716.

Tajfel, Henri, Michael G Billig, Robert P Bundy, and Claude Flament. 1971. "Social Categorization and Intergroup Behavior." *European Journal of Social Psychology* 1 (2): 149–178.

Taylor, Shelley E. 2006. "Tend and Befriend: Biobehavioral Bases of Affiliation Under Stress." *Current Directions in Psychological Science* 15 (6): 273–277.

Taylor, Shelley E, Laura Cousino Klein, Brian P Lewis, Tara L Gruenewald, Regan AR Gurung, and John A Updegraff. 2000. "Biobehavioral responses to stress in females: tend-and-befriend, not fight-or-flight." *Psychological review* 107 (3): 411.

Vincent, Carole. 2013. *Why Do Women Earn Less than Men ? A Synthesis of Findings from Canadian Microdata.* CRDCN Synthesis Series. Hamilton: Canadian Research Data Centre Network.

von Dawans, Bernadette, Clemens Kirschbaum, and Markus Heinrichs. 2011. "The Trier Social Stress Test for Groups (TSST-G): A new research tool for controlled simultaneous social stress exposure in a group format." *Psychoneuroendocrinology* 36 (4): 514–522.

Voors, Maarten J, Eleonora E M Nillesen, Philip Verwimp, Erwin H Bulte, Robert Lensink, and Daan P Van Van Soest. 2012. "Violent Conflict and Behavior: A Field Experiment in Burundi." *American Economic Review* 102 (2): 941–964.

Wildschut, Tim, Brad Pinter, Jack L Vevea, Chester a Insko, and John Schopler. 2003. "Beyond the group mind: a quantitative review of the interindividual-intergroup discontinuity effect." *Psychological bulletin* 129 (5): 698–722.

Willinger, Marc, Claudia Keser, Christopher Lohmann, and Jean-Claude Usunier. 2003. "A comparison of trust and reciprocity between France and Germany: Experimental investigation based on the investment game." *Journal of Economic Psychology* 24 (4): 447–466.

Wüst, Stefan, Ilona S. Federenko, Elisabeth F C van Rossum, Jan W Koper, and Dirk H. Hellhammer. 2005. "Habituation of cortisol responses to repeated psychosocial stress-further characterization and impact of genetic factors." *Psychoneuroendocrinology* 30 (2): 199–211.

Yanagizawa-Drott, David. 2014. "Propaganda and Conflict: Evidence from the Rwandan Genocide." *Quarterly Journal of Economics* 129 (4): 1947–1994.

Yinger, John. 1998. "Evidence on Discrimination in Consumer Markets." *Journal of Economic Perspectives* 12 (2): 23–40.