Grimm, Stefan and Klimm, Felix:

Blaming the Refugees? Experimental Evidence on Responsibility Attribution

Munich Discussion Paper No. 2018-2

Department of Economics
University of Munich

Volkswirtschaftliche Fakultät
Ludwig-Maximilians-Universität München

Online at https://doi.org/10.5282/ubm/epub.42657

Blaming the Refugees? Experimental Evidence on Responsibility Attribution

Stefan Grimm* Felix Klimm†

March 8, 2018

Abstract

Do people blame refugees for negative events? We propose a novel experimental paradigm to measure discrimination in responsibility attribution towards Arabic refugees. Participants in the laboratory experience a positive or negative income shock, which is with equal probability caused by a random draw or another participant's performance in a real effort task. Responsibility attribution is measured by beliefs about whether the shock is due to the other participant's performance or the random draw. We find evidence for reverse discrimination: Natives attribute responsibility more favorably to refugees than to other natives. In particular, refugees are less often held responsible for negative income shocks. Moreover, natives with negative implicit associations towards Arabic names attribute responsibility less favorably to refugees than natives with positive associations. Since neither actual performance differences nor beliefs about natives' and refugees' performance can explain our finding of reverse discrimination, we rule out statistical discrimination as the driving force. We discuss explanations based on theories of self-image and identity concerns.

JEL-classification: C91, D03, D83, J15
Keywords: Refugees, discrimination, responsibility attribution

*University of Munich. Contact: stefan.grimm@econ.lmu.de
†University of Munich. Contact: felix.klimm@econ.lmu.de

We would like to thank Johannes Abeler, Ingvild Almas, Yan Chen, Elwyn Davies, Lorenz Götte, Martin Kocher, Johannes Maier, Juanjuan Meng, Klaus Schmidt, and Stefan Trautmann as well as participants at seminars at the University of Munich, 2017 ESA World Meeting, 2017 ESA European Meeting and the 12th Nordic Conference on Behavioral and Experimental Economics for helpful comments. Reem Hassan and Salma Nosseir provided excellent research assistance. Ethical approval for the experiment was obtained from the Ethics Commission of the Department of Economics, University of Munich. Both authors gratefully acknowledge the financial support of the German Research Foundation (DFG) through GRK 1928 and CRC TRR 190.

"You know what a disaster this massive immigration has been to Germany and the people of Germany — crime has risen to levels that no one thought they would ever see."
U.S. president Donald Trump on refugees in Germany[1]

1 Introduction

Europe experienced a large inflow of refugees in 2015. As a consequence, a heated debate about whether to tolerate large refugee inflows or whether to instead close borders arose in both the U.S. and Europe. As reflected by the quote of U.S. president Donald Trump at the beginning of the paper, this discussion focuses to a large extent on whether refugees are responsible for negative outcomes such as rising crime rates, adverse aggregate employment, or poor economic development. Some suggest such responsibility, while others argue against it and accuse their opponents of xenophobic attitudes.[2] Despite the relevance of discrimination against refugees for social and economic outcomes, surprisingly little is known about whether natives indeed blame refugees for undesired events, and if so, whether this is caused by statistical discrimination.

We address these questions by implementing a laboratory experiment with refugees who are placed in Munich, Germany. German participants are randomly paired either with another German or a refugee. This allows us to provide clean evidence on differences in responsibility attribution and to shed light on mechanisms of discrimination in this context. More precisely, our subjects receive a positive or a negative income shock. This shock is either due to a random draw or the partner's performance in a real effort task, which took place before the main part of the experiment. If the partner actually is responsible for the shock — unbeknownst to the participant — and his performance was high enough to pass a certain threshold, a positive income shock occurs. In contrast, low performance implies a negative shock

[1] https://www.washingtonpost.com/news/worldviews/wp/2016/08/16/trump-says-german-crime-levels-have-risen-and-refugees-are-to-blame-not-exactly (last accessed on March 8, 2018).

[2] Besides the article in The Washington Post referred to in footnote 1, see https://www.nytimes.com/2016/12/09/world/europe/refugees-arrest-turns-a-crime-into-national-news-and-debate-in-germany.html (last accessed on March 8, 2018).

when the partner is responsible. After displaying the individual income shocks to the participants, we elicit beliefs about responsibility, i.e., whether the matched partner or the random draw was responsible — our core outcome measure. To investigate whether our results are driven by statistical discrimination, we further elicit beliefs about the partner's performance.[3]

This setup closely relates to many situations in which responsibility has to be assigned while there is uncertainty with respect to the actual cause. Consider, for example, employee evaluations. Increasing or decreasing sales can arise directly from the performance of an employee or be due to general shifts in demand. Layoff or promotion as well as bonus and raise decisions will crucially depend on the supervisor's assessment of this responsibility. However, responsibility attribution is not only essential for an individual's success once in a certain position, it can also critically affect the chances of being hired in the first place. The interpretation of a vita's quality signals — for example whether good performance evaluations refer to the individual's performance or merely to lenient HR policies — but also the assessment of late arrivals to interviews or sickness strongly affect hiring decisions. For all good and bad outcomes, many explanations for responsibility of either the candidate or "nature" are possible. Differing attribution behavior for refugees compared to natives can consequently have a major impact on refugees' labor market integration efforts. To the best of our knowledge, we are the first to investigate such discrimination in responsibility attribution, do so by inviting refugees — a highly relevant group for that matter — to the laboratory and implement a new experimental paradigm.

We do not observe discrimination against the outgroup of refugees by blaming them for negative outcomes. Quite the contrary can be inferred from our data. Refugees are treated more favorably than Germans. They are held responsible relatively more often for positive and less often for negative shocks. Actual performance differences and beliefs about the performance of Germans and refugees

[3]In the literature, the term statistical discrimination is most often used for discrimination based on actual differences in characteristics or behavior between different groups (e.g., Fershtman and Gneezy, 2001). Since our subjects have no information about average performances of Germans and refugees, we instead refer to discrimination based on (potentially inaccurate) *beliefs* about different performances as statistical discrimination.

cannot explain this difference. Hence, statistical discrimination does not explain our result of reverse discrimination. Furthermore, we measure implicit associations towards Arabic names and show that, despite our finding of reverse discrimination, Germans on average have negative implicit associations towards Arabic names. Indicating a positive relationship between implicit attitudes and explicit attribution behavior, subjects with positive implicit associations favor refugees more than subjects with negative associations. In addition, we do not find any evidence for reverse discrimination in a second experiment, in which we assign Germans to artificial in- and outgroups. This shows that our findings from the first experiment are driven by our natural outgroup of refugees and are not a result of our experimental design per se.

Discrimination affects a wide range of social and economic outcomes and comes in many forms and domains. For instance, discrimination can result in disadvantages for education and health related outcomes (e.g., Heckman, 1998; Shapiro et al., 2013; Krieger, 2014) as well as in obstacles to participate in the labor market (e.g., Goldin and Rouse, 2000; Carneiro et al., 2005; Lang and Manove, 2011). Our paper abstracts from these different domains and sheds light on a specific form of discrimination that has not been studied yet — responsibility attribution. Our design also allows us to distinguish between statistical and other types of discrimination and hence to talk about the channels for discriminatory behavior. Other experimental papers have specifically looked at a variety of underlying mechanisms, too.[4] Fershtman and Gneezy (2001) investigate trust and social preferences of ingroup and outgroup members in the Israeli society. Using the investment, dictator, and ultimatum game, they find clear stereotypes associated with different ethnic groups leading to discriminatory behavior. Ockenfels and Werner (2014) provide related evidence on ingroup favoritism. They show that people share more of their endowment in a dictator game when paired with an ingroup member, which indicates an explanation based on social preferences. Similarly, Chen and Li (2009) report increased altruism towards ingroup members in allocation games for different measures of social

[4]For a meta-study on economic experiments on discrimination, see Lane (2016).

preferences, e.g., punishment for misbehavior. In stark contrast to these papers, we do not observe ingroup favoritism or discrimination "against" the outgroup but document reverse discrimination.

We also contribute more generally to the understanding of how responsibility is attributed per se. Bartling and Fischbacher (2011) and Bartling et al. (2015) show that responsibility can be effectively shifted through the delegation of choice and not being pivotal. This evidence indicates that responsibility attribution is malleable and that there is scope for discrimination in attribution behavior.

The much more extensive literature on responsibility attribution in psychology focuses on whether individuals attribute explicit behaviors to internal characteristics or situational factors. Ross (1977) coined the term "fundamental attribution error", which presumes the tendency to underestimate the role of external circumstances when judging others' behavior. Jones and Harris (1967), the original paper to address this issue, investigate subjects' assessments of a writer's private opinion of Fidel Castro. Although subjects know that the writer was randomly told to either praise or criticize Castro in an essay, they rated the writer's opinion as more favorable towards Castro when he had written a pro-Castro text. Hence, subjects wrongfully attributed responsibility for the content of the text to the writer. Pettigrew (1979) relates this bias to ingroup favoritism and hence discriminatory behavior calling it "ultimate attribution error". Negative actions by an outgroup member will more likely be attributed to personal causes, whereas positive actions are more likely attributed to external factors (e.g., luck or "the exceptional case") compared to actions by an ingroup member (for an extensive review see Hewstone, 1990). In contrast to this literature, we do not study whether internal or external factors cause individual behavior. This would correspond, for example, to attributing responsibility for an employee's explicit action. That is, the supervisor knows that the sales manager hired an excellent sales rep but can either attribute this to excellent knowledge of human nature or to mere luck. Instead, we investigate whether an event where the true underlying cause is unknown — who hired the sales rep — is attributed to an individual or something else — the specific sales manager or someone else.

As our subjects are willing to sacrifice part of their payoffs in order not to blame refugees, our finding is not compatible with the standard economic model of purely self-interested agents. Instead, we interpret our results as being in line with theories of economics of identity and motivated beliefs. In such a framework, people care about a positive self-image or generally want to behave according to certain prescriptions pertaining to their identity (Akerlof and Kranton, 2000). These concerns can affect behavior and may lead to self-serving beliefs over behavior of other people (e.g., Di Tella et al., 2015). For our context, it is important that being open and tolerant towards minorities and refugees is part of the social identity of many people, presumably especially in our student sample. Hence, identity concerns might motivate our participants to attribute responsibility more positively towards refugees since blaming refugees is clearly associated with xenophobic attitudes.[5] We also favor this interpretation because in our anonymous laboratory setting, we rule out social image concerns as much as possible.

The remainder of the paper is structured as follows. Section 2 describes the experimental design in detail. Section 3 presents our results on responsibility attribution. Section 4 is about a robustness experiment that we ran with artificially formed groups. Section 5 discusses our main finding and Section 6 concludes.

2 Experimental Procedures and Design

2.1 Procedural Details

We programmed and conducted the experiment with "z-Tree" (Fischbacher, 2007). Germans, 152 students from various fields of study, were recruited using the online recruiting system "ORSEE" (Greiner, 2015). Additionally, 43 refugees were recruited in Munich with leaflets at refugees camps, in front of local registration offices, and in cooperation with the NGO *Social Impact Recruiting* (SIR).[6] Figure A.1 in the Appendix

[5]For instance, see `http://www.independent.co.uk/voices/justin-welby-is-wrong-it-is-racist-to-blame-migrants-for-your-fears-about-jobs-and-wages-a6925106.html` (last accessed on March 8, 2018).

[6]SIR supports refugees in finding a job by creating a German CV, preparing for interviews, and contacting employers. For further information see `http://si-recruiting.org/` (last accessed on

shows an English version of the leaflet.

Because the vast majority of SIR clients and most of the refugees arriving in Germany were male, we decided to restrict the sample to male refugees.[7] Consequently, we also invited only male Germans to have single sex pairs in both ingroups and outgroups such that we did not have to control for potential gender effects. In addition, we wanted our refugee subjects to be of roughly the same age as our other participants. Hence, only refugees between the age of 18 and 29 were invited to participate in the experiment. To have a relatively homogeneous outgroup that represents the majority of refugees in Germany, we only invited Arabic native speakers.[8] To also have a homogeneous ingroup, we only invited native participants with a German sounding name. This ensured that participants assigned to an ingroup member indeed regarded the matched participant as ingroup member.[9]

All 10 experimental sessions took place at the Munich Experimental Laboratory for Economic and Social Sciences (MELESSA) at the University of Munich from August to November 2016. The assignment to the seats in the laboratory made clear that there were two different groups in the experiment. Refugees had to draw a card with a seat number from a bag with the label "Arabic" (in Arabic letters) and Germans from a bag with the label "German" (in German). The cards ensured that the participants were seated in front of a computer screen with instructions in the respective language. Within each group, subjects were randomly assigned to a seat. An English version of the instructions is included in Appendix E. Refugees were invited to the experiment half an hour earlier than Germans to make sure they knew what to expect and to check reading and writing proficiency in Modern Standard Arabic.[10] Announcements

March 8, 2018).

[7]See page 21 of the German report of the German Federal Office for Migration and Refugees: `http://www.bamf.de/SharedDocs/Anlagen/DE/Publikationen/Broschueren/bundesamt-in-zahlen-2015.html` (last accessed on March 8, 2018).

[8]German Federal Office for Migration and Refugees: `http://www.bamf.de/SharedDocs/Anlagen/EN/Publikationen/Migrationsberichte/migrationsbericht-2015-zentrale-ergebnisse` (last accessed on March 8, 2018).

[9]All refugees indeed had Arabic names. See Section A in the Appendix for a complete list of first names of all participants. At the time of writing this paper, only roughly 3% of our regular subjects registered for experiments at the Munich Experimental Laboratory for Economic and Social Sciences (MELESSA) had Arabic sounding names. It therefore should have been clear to our German participants that they were matched with a refugee when their partner's name was Arabic sounding.

[10]Some refugees could not participate in the experiment since they indicated that they were not

before and during the experiment were repeated in Arabic by two student research assistants. If necessary, they answered questions by the refugees individually at the subjects' seats. Questions of Germans were answered by the experimenter.

For the main part of the experiment, we formed ingroup and outgroup pairs. As we do not focus on how refugees attribute responsibility, we denote Germans matched with another German as belonging to the *German* treatment (ingroup) and Germans matched with a refugee as belonging to the *Refugee* treatment (outgroup). In order to increase the number of decisions taken by Germans, we matched each refugee with up to two Germans. Group assignment of Germans was random conditional on assigning the same number of Germans to the treatments *German* and *Refugee*.[11] At the beginning of the main part of the experiment, subjects needed to enter their first name, which was then shown to their matched partner and enabled all subjects to identify their partner's group affiliation.[12]

At the end of the experiment, the participants answered a questionnaire about socio-demographic characteristics. Thereafter, all subjects were paid privately and earned €12.3 on average, including a fixed payment of €6 for showing up on time. The sessions lasted between 60 and 75 minutes. Each subject participated in one session only.

2.2 Experimental Design

Our experiment consisted of two parts. In the first part, subjects received a flat fee of €3 for performing a real effort task. They solved up to eight simple (6×4) jigsaw puzzles (henceforth puzzles) within ten minutes. The puzzles were placed next to the keyboard and were covered by a sheet of paper at every seat. Subjects were asked not to touch the stack until the experimenter had indicated to begin. We chose puzzle

sufficiently able to read and spell.

[11]Only even numbers of German subjects participated in the sessions. If dividing the number of German subjects into two groups of equal size resulted in an odd number, groups were formed such that there were two more Germans matched with a refugee than with another German. For instance, in a session with 18 Germans, 10 of them were matched with a refugee.

[12]Loss of anonymity is not a concern despite identification via names. In the questionnaire at the end of the experiment, only 6% of German participants indicated that they knew another participant in their session. There are on average more than 15 German participants per session. Hence, their likelihood of being matched to someone known is smaller than 1%.

motives to be culturally neutral (see Figure B.1 in the Appendix). This real effort task has the advantage of being familiar to participants from different parts of the world. We could not use a computer-based task because many of the refugees were not familiar with working with a personal computer.[13] Furthermore, many Germans arguably would have expected a large performance difference between refugees and Germans. Importantly, at the time of solving the puzzles, participants knew nothing about the content of the rest of the experiment. At the end of part one, the experimenter and student research assistants quietly counted the number of correctly solved puzzles at the subjects' seats.

For the second and main part of the experiment, subjects were randomly paired with another participant in the experiment into ingroup (both subjects Germans) and outgroup pairs (one German and refugee each). The decision task of the second part of the experiment is illustrated in Figure 1. Player A faced a positive or negative income shock. He either received €5 or €5 were subtracted from his experimental earnings.[14] However, player A did not know how this shock came about. With an ex-ante probability of 50%, this shock was due to the performance of player B (the matched participant) and otherwise due to nature. If player B's performance was responsible for the income shock, the shock was positive if player B's number of correctly solved puzzles was at least four and negative otherwise. In the case of nature being responsible for the income shock, one of the two shocks was randomly chosen with equal probability. Furthermore, player B's payoff was not affected by whether player A received a positive or negative shock.

The decision task was performed symmetrically within each pair, i.e., every subject was player A and player B. Subjects were fully aware of the task setup. All participants had to answer four control questions correctly before starting the main part of the experiment to make sure they fully understood the decision tree.

Subsequently, in the first belief elicitation, subjects guessed whether nature or player B's performance caused the income shock and received €5 if their guess was

[13] In the first three sessions, we asked refugees whether they are familiar with puzzles before the start of the experiment. All of them confirmed.

[14] Subjects knew that their total earnings from the experiment would be a positive amount.

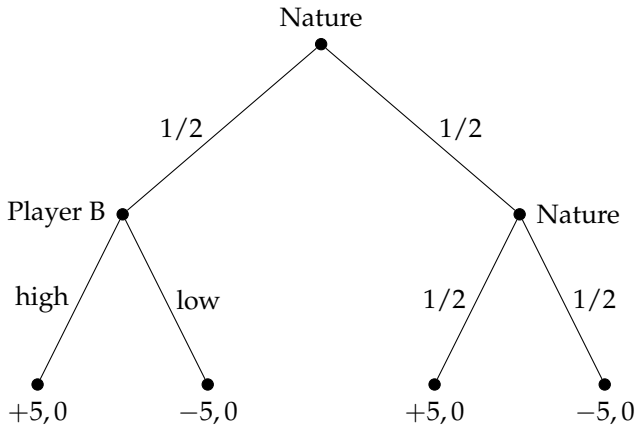


Figure 1: Decision tree

correct. This allows us to identify differences in responsibility attribution to Germans and refugees and is our main variable of interest. In order to get a more precise measure of responsibility attribution, we additionally asked for the player's confidence in their own guess in a second belief elicitation. More specifically, participants filled out a 9-item choice list with two options (A and B) for each of the nine choices (based on Becker et al., 1964, henceforth BDM). If they chose option A and the respective choice became payoff relevant, they received €5 if their chosen mechanism (in the first belief elicitation) was indeed responsible for the shock (player B or nature). Option A was the same for all nine choices. Option B gave them the chance to receive €5 with probabilities ranging from 10% to 90% in 10% increments. If a participant, for example, expected player B to be responsible in the first elicitation and switched to option B in row seven, he assigned between 60% and 70% probability to the event that player B indeed was responsible.

In addition, we elicited binary beliefs about performance to see whether potential differences in responsibility attribution stem from statistical discrimination. We asked whether subjects believed that the matched player's performance passed the threshold of four solved puzzles or not (again incentivized with €5). Finally, we asked for the probability player A assigned to the matched participant having solved at least four puzzles. Again, subjects faced a (BDM-based) choice list with nine choices between option A, i.e., receiving €5 if the partner's performance was at or above the cutoff, and option B, i.e., receiving €5 with given probabilities ranging from 10% to 90%. Hence,

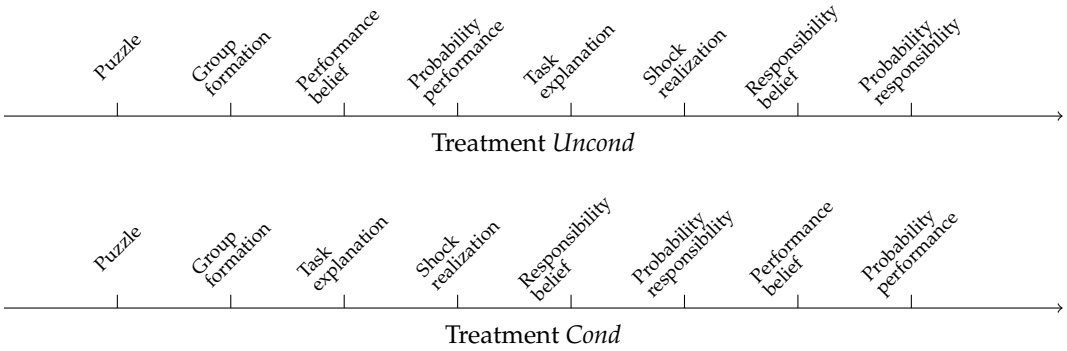


Figure 2: Timeline of the experiment

in total, we elicited four incentivized beliefs. At the end of the experiment, in order to prevent hedging, one of these belief questions was randomly chosen for payment and either paid €5 or nothing.

The order of the four belief elicitations, however, was not the same in all sessions. In half of the sessions, we elicited performance beliefs before explaining the structure of the decision task. Hence, in these sessions (henceforth *Uncond*), participants first worked on the puzzles, were then matched with a partner and directly asked for the two (unconditional) performance beliefs regarding the partner (binary choice and choice list). Only then the decision task was explained and the shock realized. In the other half of the sessions (henceforth *Cond*), (conditional) performance beliefs were elicited after the task had been explained, the shock had realized, and after subjects had attributed responsibility. This allows us — by comparing performance beliefs in the treatments *Uncond* and *Cond* — to examine whether subjects formed distorted or motivated beliefs after observing the shock and attributing responsibility. For instance, assume that a subject attributes responsibility to the partner after observing a negative shock. If this subject is asked about his performance belief, he could justify his attribution behavior by stating low performance beliefs, although he actually thinks that the partner passed the cutoff. Hence, we had a 2×2 treatment design along the dimensions group assignment and task order. Figure 2 provides an overview of task orders in the respective treatments.

After these two main parts of the experiment, participants performed the Implicit Association Test (IAT) to measure implicit associations towards Arabic names. Subjects

had to assign positive (e.g., "appealing", "love", "cheer") or negative expressions (e.g., "selfish", "dirty", "bothersome") to Arabic or Caucasian names by pressing keys on their keyboard. The IAT score, which indicates positive or negative associations towards Arabic names, is calculated based on response times to sort names to expressions. If a subject needed more time to assign positive expressions and less to assign negative expressions to Arabic compared to Caucasian names, the IAT score is below zero indicating negative implicit attitudes towards Arabic names. This task has been shown to relate to various dimensions of field behavior such as job recruitment (see Greenwald et al. (2009) for a meta study). We used FreeIAT, a free software to run IATs.[15] Subjects were paid €2 for completing the IAT.

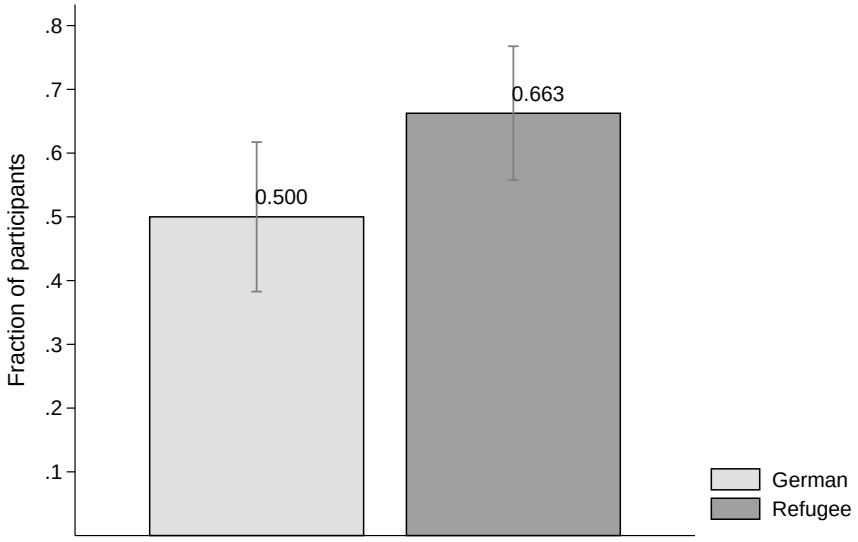
3 Results

Our main results on the comparison of responsibility attribution by group assignment over all sessions combined are reported in Section 3.1. This abstracts from potential systematic differences between *Uncond* and *Cond*, which we analyze in 3.2 separately. Section 3.3 presents evidence for heterogeneity using scores from the Implicit Association Test. Section 3.4 reports results using the BDM-based probability measures of our main outcome variable and performance beliefs. Unless stated otherwise, all our results in this section consider attribution behavior of our German participants only.

3.1 Favorable Responsibility Attribution

Since we test whether our subjects assign responsibility less, equally or more favorably to Germans or refugees, i.e., whether there is discrimination in attribution behavior, we define the binary variable *favorable attribution*. We denote responsibility attribution as favorable if a positive shock occurs and the matched partner is believed to be responsible for the shock. Attribution is also favorable if a negative shock is observed and responsibility is assigned to nature. In contrast, attributing responsibility to the matched partner after a negative shock or to nature after a positive shock

[15] `http://www4.ncsu.edu/~awmeade/FreeIAT/FreeIAT.htm` (last accessed on March 8, 2018).



Notes: The figure shows *favorable attribution* for both treatments. Error bars indicate 95% confidence intervals.

Figure 3: *Favorable attribution* depending on group affiliation

implies unfavorable attribution.[16] This simplification ignores potential asymmetries in behavior after positive versus negative income shocks. We will show later that our results hold for both shock directions.

Figure 3 displays *favorable attribution* by group affiliation. Germans matched with another German ($n = 72$) equally often attribute responsibility favorably and unfavorably. In stark contrast to that, Germans matched with a refugee ($n = 80$) attribute responsibility favorably in roughly two thirds of the cases. This difference in attribution behavior is statistically significant ($p = 0.042$, χ^2-test, two-sided) and evidence for reverse discrimination, i.e., a positive bias towards the refugee outgroup.

Under rationality, *favorable attribution* represents the belief about the matched partner having solved at least four puzzles. Hence, the results displayed in Figure 3 could be driven by performance beliefs depending on group affiliation. We would expect more favorable attribution in *Refugee* if subjects believed that refugees are better than Germans in solving puzzles. However, comparing performance beliefs reveals no

[16]The intuition underlying this distinction is rational behavior depending on beliefs. Nature and the matched partner are ex-ante responsible with equal probability. Given nature is responsible, positive and negative shocks occur with equal probability. Hence, if the decision maker expects the matched partner to having solved four or more puzzles and thus assigns a probability larger than 50% to this event, he should attribute responsibility favorably. Therefore, under the assumption of rational behavior, *favorable attribution* captures underlying beliefs about the partner reaching the puzzle cutoff.

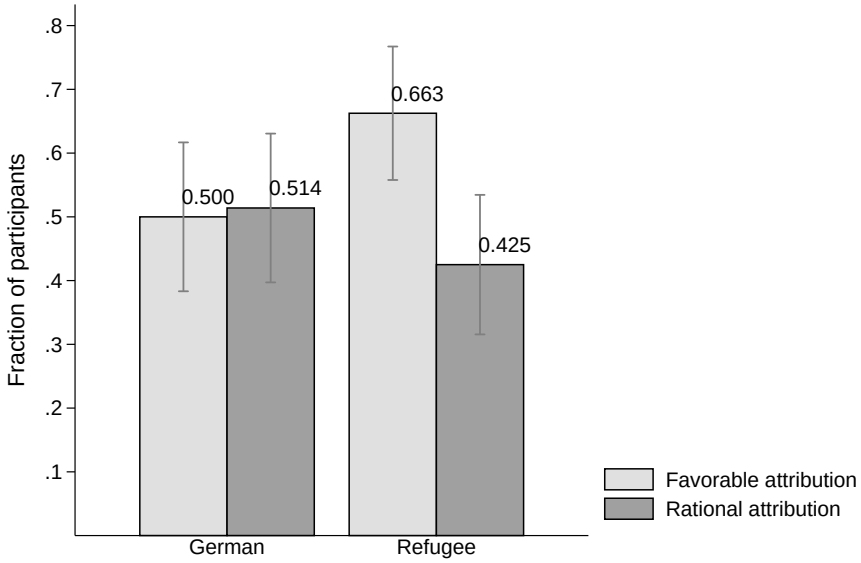
significant difference. If anything, Germans expect refugees to perform slightly worse, which renders reverse discrimination even more pronounced. While 43% of Germans matched with a refugee expect the refugee to have solved at least four puzzles, 51% of Germans matched with another German have high performance beliefs ($p = 0.273$, χ^2-test, two-sided).[17] This indicates that the asymmetry in responsibility attribution cannot be rationally based on performance beliefs. In Figure 4, we compare actual favorable responsibility attribution (*favorable attribution*) and rational favorable responsibility attribution (*rational attribution*). We define *rational attribution* to be one if the German participant has high performance beliefs regarding the matched partner and zero otherwise. Figure 4 shows that while actual responsibility attribution is on average in line with performance beliefs for Germans matched with another German, attribution is clearly more favorable than dictated by performance beliefs for Germans matched with refugees.[18] The difference in *Refugee* is significant ($p < 0.01$, McNemar test, two-sided).[19]

Next, we control for the direction of the income shock. Since the actual performance of refugees was much worse than that of Germans, Germans in *Refugee* observe negative shocks much more often. Hence, more favorable attribution after negative shocks, independent of group affiliation, could explain our results. However, the shock direction does not drive our finding. For both negative and positive shocks, there is a clear asymmetry by group affiliation in terms of how performance beliefs translate into responsibility attribution (see Figure C.1 in the Appendix). Importantly, there is no evidence for blaming the refugees in case of negative shocks. We observe the contrary. Refugees are attributed responsibility much more favorably after a negative

[17] With our sample size, we have 80% power to detect an effect size that implies a belief difference of around 22 percentage points. Actual performance differences are much more pronounced. While 47% of the Germans solve four or more puzzles, only 2.3% of the refugees (1 out of 43) reached the performance cutoff. Therefore, statistical discrimination based on actual behavior would imply much more favorable attribution to Germans and thus can neither explain our results.

[18] We cannot analyze refugee behavior by group affiliation since refugees are only matched with Germans. While this is not the interest of this paper and we do not have adequate power to detect patterns, 51.2% attribute responsibility favorably, whereas only 9.3% of them believe that their partner made the performance cutoff.

[19] These findings are robust to comparing attribution behavior with the individual's own performance. While own performance need not necessarily be a perfect proxy for beliefs regarding the performance of the other, performance is certainly orthogonal to treatment — unlike beliefs that could potentially be affected by treatment. We will extensively discuss this in Section 3.2.



Notes: The figure shows *favorable attribution* and *rational attribution* implied by beliefs for both treatments. Error bars indicate 95% confidence intervals.

Figure 4: *Favorable attribution* and *rational attribution* implied by beliefs

shock compared to rational attribution based on performance beliefs ($p < 0.01$, McNemar test, two-sided).

To verify the robustness of our non-parametric results, we run different regression models. The regression framework helps us to further understand attribution behavior by explicitly measuring the effects of beliefs and shock direction on *favorable attribution* while being able to control for observables, too. Table 1 reports marginal effects from probit regressions on our binary variable *favorable attribution*.

Column (1) is the parametric equivalent to Figure 3 replicating the significant positive effect of being matched with a refugee on *favorable attribution*. This is indicated by the binary variable *Refugee*, which is equal to one if a subject is matched with a refugee and zero otherwise. Column (2), equivalent to Figure 4, controls for performance beliefs with *belief high* as binary variable. *Belief high* is equal to one if a subject believes that the partner passed the cutoff and zero otherwise. The effect of group affiliation remains highly significant and sizable. Being matched with a refugee increases the likelihood to attribute responsibility favorably by 19.5 percentage points. The effect in model (2) is slightly larger than in model (1), which is in line with our non-parametric results. As performance beliefs are slightly worse for refugees, controlling for beliefs increases the effect of group affiliation. Reassuringly, high

Table 1: Favorable responsibility attribution

Dependent variable	Favorable attribution			
	(1)	(2)	(3)	(4)
Refugee	0.160***	0.195***	0.155***	0.146***
	(0.056)	(0.050)	(0.040)	(0.038)
Belief high		0.372***	0.369***	0.375***
		(0.067)	(0.070)	(0.068)
Neg shock			0.164**	0.158**
			(0.064)	(0.064)
Additional controls	No	No	No	Yes
Observations	152	152	152	152
Pseudo R^2	0.020	0.149	0.172	0.179

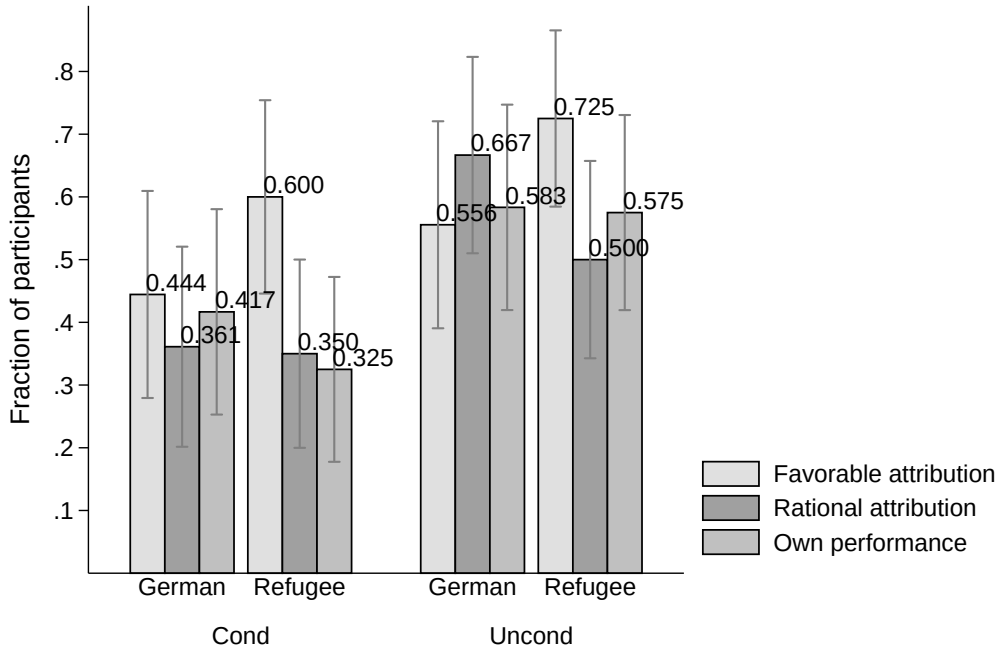
Notes: Probit regressions on *favorable attribution* reporting average marginal effects. Column (4) includes additional covariates from the questionnaire: age, semester, and number of experiments so far (all insignificant). Robust and clustered (on session level) standard errors in parentheses. Stars indicate significance on the levels: *** p<0.01, ** p<0.05, * p<0.1.

performance beliefs lead to more favorable responsibility attribution. Subjects who believe that the partner passed the cutoff are 37.2 percentage points more likely to exhibit favorable attribution. As motivated above, we include the shock direction in column (3) with *neg shock* as binary variable. It is equal to one if a negative shock occurs and zero otherwise. We find a significant positive effect of negative shocks indicating that participants attribute responsibility generally more favorably after a negative shock. However, this does not alter our finding regarding group affiliation. Finally, our results are robust to controlling for personal background variables in column (4).

Result 1: *Germans attribute responsibility more favorably to refugees than to other German participants. This cannot be explained by differing performance beliefs and holds for behavior after both negative and positive shocks.*

3.2 Unconditional vs. Conditional Beliefs

Participants in our *Cond* treatment were asked to state their performance beliefs after observing the shock and after attributing responsibility. Hence, in order to justify attribution in front of themselves, participants may report distorted beliefs. To quantify this potential distortion, we ran half of the sessions with performance beliefs elicited before shock realization and responsibility attribution (*Uncond*).



Notes: The figure shows *favorable attribution*, *rational attribution*, and the fraction of participants reaching the puzzle cutoff (*own performance*) by group affiliation for the treatments *Cond* (left panel) and *Uncond* (right panel). Error bars indicate 95% confidence intervals.

Figure 5: *Favorable attribution*, *rational attribution*, and *own performance*

To investigate whether performance beliefs are distorted, we relate these beliefs to *own performance* — measured by whether the individual solved at least four puzzles. *Own performance* serves as a benchmark for beliefs regarding others' performances and hence should be the main driver for performance beliefs. This hypothesis is supported by our data. In *German*, 50% pass the puzzle cutoff and 51% expect the matched partner to having done so. In *Refugee*, 45% of Germans solve at least four puzzles and 43% expect that from the matched partner. Only roughly one fourth of our subjects, both in *German* and *Refugee*, does not believe the matched participant to have performed in the same way as they did. Figure 5 displays average *own performance*, beliefs in the other's performance (i.e., *rational attribution*), and actual responsibility attribution (*favorable attribution*) by group affiliation and task ordering (*Uncond* vs. *Cond*) separately.[20]

Performance beliefs cannot be distorted by knowledge about our responsibility attribution task in *Uncond*. In this case, displayed in the right panel of Figure 5,

[20]This reveals that randomization was not successful with regard to puzzle performance. A significantly larger fraction of subjects in *Uncond* pass the performance cutoff than subjects in *Cond* ($p < 0.01$, χ^2-test, two-sided). Table C.1 in the Appendix shows the sample balance.

Germans expect other Germans on average to perform slightly better than themselves and refugees to be slightly worse. Compared to that, performance beliefs seem distorted in *Cond*. Beliefs of ingroup members are slightly lower than *own performance*, while they are higher for Germans in *Refugee*. On average, Germans matched with a refugee in *Uncond* are 7.5 percentage points less likely to believe in the performance of their partner compared to their own performance. However, German outgroup participants in *Cond* are 2.5 percentage points more likely to believe in the performance of the refugee than in their own. Hence, the difference in the differences between *own performance* and performance beliefs over the two treatments for subjects in *Refugee* is 0.1. This corresponds to a positive belief distortion in favor of refugees once knowing the decision task. Performing the same difference in differences calculation for subjects in *German*, we find a difference in differences of 0.14 that shows worse performance beliefs in *Cond* (negative distortion against other Germans). While this 24 percentage points difference in distortion between *German* and *Refugee* is considerate, it is insignificant ($p = 0.151$, t-test, two-sided).[21]

Hence, under the assumption of unbiased beliefs in *Uncond* our findings from Section 3.1 provide a lower bound for the extent of reverse discrimination. The results from this section indicate that true underlying beliefs in *Cond* could actually be worse for refugees and better for other Germans than stated in the belief elicitation. This would increase the asymmetry between rational and actual responsibility attribution beyond what we measure in Section 3.1.

Result 2: *We find no significant evidence for subjects stating distorted beliefs. However, if anything, the results point towards favorably distorted beliefs with respect to refugees, suggesting that the results from the pooled sample (Section 3.1) constitute a lower bound for reverse discrimination.*

The assumption in this section is that beliefs in *Uncond* are unbiased. This seems reasonable since participants are unaware of the rest of the experiment in this treatment

[21] This calculation is equivalent to regressing the individual difference between *rational attribution* (performance beliefs) and *own performance* in an OLS estimation on *Refugee*, *Cond*, and their interaction term *Refugee*×*Cond*. The interaction term shows the 24 percentage points distortion for Germans matched to refugees once they know the decision task.

when stating their guess about their partner's performance. However, unconditional performance beliefs regarding refugees could already be distorted upwards such that true underlying performance beliefs would actually be lower. If this was the case, our overall finding of reverse discrimination would again be a lower bound of the true discrimination. Given true performance beliefs, the difference between these beliefs and responsibility attribution would be larger than the one we find with stated beliefs. In contrast to that, performance beliefs could also be biased downwards and explain our result of reverse discrimination. This, however, seems very unlikely because it would imply discrimination at the level of performance beliefs — by stating lower than actual beliefs about performance for refugees — and, to the contrary, reverse discrimination at the level of responsibility attribution. Furthermore, it is implausible that participants have such extremely inaccurate beliefs given that refugees actually perform very poorly in the real effort task.

To account for the possibility of biased performance beliefs, we substitute these beliefs by own performance to check the robustness of our main findings. Table C.2 in the Appendix reports results from regressions replicating Table 1 while using each participant's number of correctly solved puzzles as explanatory variable instead of his performance beliefs.[22] The results for *Refugee* from all models are strikingly similar to the ones from Table 1, which renders our finding of reverse discrimination robust to belief distortions.

3.3 Implicit Associations

The key personal characteristic that we elicit and correlate with attribution behavior relates to implicit associations. The IAT measures people's relative implicit associations towards a specific group compared to a baseline group. In our case, it is a measure of associations towards Arabic names relative to Caucasian names.[23] A positive test score

[22] Alternatively, using a binary variable for whether the respective participant solved at least four puzzles does not change the significance of the *Refugee* or *neg shock* indicators.

[23] Arabic names are Hakim, Sharif, Yousef, Wahib, Akbar, Muhsin, Salim, Karim, Habib, and Ashraf, and Caucasian Names are Ernesto, Matthais, Maarten, Philippe, Guillame, Benoit, Takuya, Kazuki, Chaiyo, and Marcelo. Positive associations are Excellent, Cheer, Delight, Joyous, Excitement, Cherish, Friendship, and Beautiful, and negative associations are Hate, Pain, Gross, Failure, Rotten, Humiliate, Sickening, and Horrible. The IAT for Arabic names can be taken online by visiting `https://implicit.`

implies relatively positive associations towards Arabic names, while a negative score indicates the opposite.

Overall, the results from the IAT are in line with ingroup favoritism. While 72% of Germans have a negative IAT and hence relatively more negative associations towards Arabic names, this is the case for only 12% of the refugees ($p < 0.01$, χ^2-test, two-sided).[24]

Importantly, implicit attitudes have predictive power for explicit discrimination behavior. People with negative IAT scores favor refugees less with regard to responsibility attribution. 83% of Germans with a positive IAT in *Refugee* attribute responsibility favorably, while only 59% with a negative IAT do so. This difference is significant ($p = 0.034$, χ^2-test, two-sided).

To test the correlation between implicit associations and *favorable attribution* when holding other variables constant, we further apply a regression framework. We control for own performance rather than for performance beliefs since beliefs might have been distorted, and this potential distortion is likely to be related to the IAT score. For instance, subjects who are in general favorable towards refugees are likely to have a positive IAT score *and* possibly upwards biased beliefs about a refugee's performance.

Table 2 reports probit regressions of *favorable attribution* on $IATneg$, which is equal to one if the IAT score is negative (negative associations towards Arabic names) and zero otherwise (positive associations towards Arabic names), and own performance. Column (1) includes subjects in *Refugee* only. As indicated by our non-parametric results discussed before, we observe a large and significant correlation between having a negative IAT score and responsibility attribution for Germans matched with refugees. Those that have negative implicit association towards Arabic names are 27.2 percentage points less likely to attribute responsibility favorably to their matched Arabic partner. Column (2) shows that a negative IAT score has no effect on favorable responsibility attribution in *German*.[25] Column (3) reports regression results for the

harvard.edu/implicit/selectatest.html and selecting "Arab-Muslim IAT".

[24]The same holds true for average values. The average IAT score for Germans is -0.199, while the average for refugees is 0.215. This difference is again highly significant ($p < 0.01$, Mann-Whitney U-test, two-sided).

[25]Ex-ante, it is not obvious why the effect of implicit associations should be stronger in *Refugee*

entire sample with additional controls and an interaction of the IAT score and our treatment. These results confirm our findings from column (1) and (2). The marginal effect of the interaction term of –0.343 indicates that a negative IAT value has a more negative effect on *favorable attribution* for participants in *Refugee* compared to participants in *German*. Further, replicating columns (1) and (2), we see that IAT scores (*IATneg*) do not affect *favorable attribution* in *German*. In contrast, having a negative IAT score decreases the likelihood to attribute responsibility favorably by 25.9 percentage points in *Refugee* ($p = 0.030$, *F*-Test for *IATneg* + *IATneg* x *Refugee*).[26] In addition, the coefficient of *Refugee* shows that our result of reverse discrimination is mainly driven by participants with a positive IAT score since the treatment difference is insignificant for subjects with a negative IAT score ($p = 0.390$, *F*-Test for *Refugee* + *IATneg* x *Refugee*).

However, in nonlinear models including interaction terms, interpreting the marginal effect of the interaction term is flawed (Ai and Norton, 2003) and hypothesis testing can be misleading (Greene, 2010). This is due to the fact that, in nonlinear models, the marginal effect of the interaction term is not the same as the cross derivative with respect to both interacted variables (the interaction effect). In order to account for this problem, we compute the predicted values of *favorable attribution* split up along two dimensions — having a positive or negative IAT score as well as being in *Refugee* or *German*. We calculate the difference in differences of these four groups, which reflects the interaction effect in models including interaction terms with two binary variables. We find that the effect of a negative IAT score on *favorable attribution* is 36.19 percentage points lower in *Refugee* than in *German*.[27] Since this estimate is very close to the marginal effect of our interaction term in column (3), –0.343, the mistake

compared to *German*. The effects in the two different groups should go into opposite directions, but there is no apparent reason why positive implicit associations towards one's ingroup should not lead to more favorable attribution towards these ingroup members. We interpret this finding in the following way. First, it is plausible that associations regarding the more salient outgroup determine the IAT scores. In that case, the IAT score should not predict behavior towards the ingroup. Second, we used a standard version of the IAT measuring associations towards Arabic names. This version uses a wide range of Caucasian names in the baseline group. Hence, attitudes towards German participants might not be perfectly captured by this IAT. This again supports the idea that our IAT scores predominantly represent implicit associations towards Arabic names and not German names.

[26] All results from Table 2 are qualitatively unchanged if we use the continuous variable of the IAT instead of the binary version. Only the *F*-Test for *IAT* + *IAT* x *Refugee* in the interaction model becomes borderline insignificant ($p = 0.143$).

[27] Estimation of the difference in differences in predicted values can be found in Appendix D.

Table 2: Favorable responsibility attribution depending on IAT

Dependent variable	Favorable attribution		
	Refugee (1)	*German* (2)	pooled (3)
IATneg	−0.272**	0.089	0.084
	(0.114)	(0.159)	(0.162)
# correct puzzles	0.077**	0.104***	0.092***
	(0.036)	(0.024)	(0.020)
Refugee			0.395***
			(0.146)
IATneg × Refugee			−0.343*
			(0.186)
Neg shock			0.123**
			(0.058)
Additional controls	No	No	Yes
Observations	80	72	152
Pseudo R^2	0.076	0.071	0.114

Notes: Probit regressions on *favorable attribution* reporting average marginal effects. Column (1) and (2) include only the sample of outgroup and ingroup participants respectively. Column (3) includes the entire sample and additional covariates from the questionnaire: age, semester, and number of experiments so far (all insignificant). Robust and clustered (on session level) standard errors in parentheses. Stars indicate significance on the levels: *** p<0.01, ** p<0.05, * p<0.1.

induced by interpreting the marginal effect of the interaction term as interaction effect is negligible in our estimation.

Result 3: *Implicit associations directly relate to explicit behavior. Reverse discrimination is mainly driven by subjects with positive implicit association towards Arabic names.*

3.4 Alternative Measures of Responsibility Attribution and Performance Belief

By using the binary measure of responsibility attribution and by enforcing a choice, we treat more or less indifferent participants the same as those who have a clear opinion about responsibility. In this section, we want to check whether these indifferent people could be driving our results. For this purpose, we define two new variables called (i) *responsibility switchpoint* and (ii) *performance switchpoint* based on the two BDM belief elicitations. These variables indicate probabilistic confidence in (i) the partner being responsible for a positive shock (conditional on observing a positive shock) or the

Table 3: Contingency table for binary vs. BDM choices

	1	2	3	4	5	6	7	8	9	10
Responsibility:										
(1) Binary favorable: Switchpoint	0	2	0	3	21	31	18	11	2	1
(2) Binary unfavorable: Switchpoint	3	2	7	14	16	14	2	4	1	0
Performance:										
(3) Binary positive: Switchpoint	0	0	0	3	10	14	23	12	7	2
(4) Binary negative: Switchpoint	3	5	12	21	22	11	2	2	3	0

partner *not* being responsible for a negative shock (conditional on a negative shock) and (ii) the partner having solved four or more puzzles. A higher value of *responsibility switchpoint* hence indicates a more favorable attribution. A higher value of *performance switchpoint* indicates a higher confidence in the matched partner having solved four or more puzzles. Both variables, corresponding to the nine-item choice list, are measured in 10 percentage point steps. Thus, a switchpoint of one corresponds to assigning 0-10% probability to the event and a switchpoint of 10 corresponds to 90-100%.

The average of *responsibility switchpoint* by group affiliation highlights a clear difference to the findings from the binary measure. With an average switchpoint of 5.65 and 5.56 in *German* and *Refugee* respectively, there is no difference in responsibility attribution by group affiliation. Is this difference in response behavior driven by outliers, by indifferent participants, or do we observe other inconsistencies? To understand consistency between the binary and BDM belief elicitation, Table 3 displays a contingency table for these choices reporting combinations of binary choices and BDM choices. Row (1) and (2) refer to responsibility consistency, given that in the binary choice responsibility was assigned favorably (1) or unfavorably (2). Rows (3) and (4) display consistency for performance beliefs depending on the binary performance belief elicitation.

If consistent, row (1) subjects should have a *responsibility switchpoint* above five and thus assign more than 50% probability to the "favorable" event. Those around the threshold are close to indifference (highlighted in dark gray), while those in light gray choose clearly inconsistently. For instance, assigning only 30-40% probability to the matched partner being responsible for a positive shock but before indicating to

believe the partner is responsible — as is the case for the three participants highlighted in row (1) in the fourth column — is not consistent. The table shows that a substantial fraction of participants reports probabilities around the indifference threshold of 5 and 6, indicating that indifference could help to explain our difference in non-parametric results between our binary and BDM responsibility measures.

Moreover, it seems that some subjects did not understand the BDM choice list. Twelve participants strongly violate consistency when asked about responsibility, and ten participants do so for the performance beliefs. In line with the notion of misunderstanding, it takes these participants also clearly longer to make these BDM choices. Those being inconsistent for the performance questions take on average 24 seconds longer (out of 90 seconds they have) for this BDM, while they are 2.5 seconds faster than the consistent subjects for the binary performance belief (both comparisons do not exceed a p-value of 0.037, Mann-Whitney U-test, two-sided). Directionally, the same is true for the responsibility questions. Participants that are inconsistent spend on average 3.5 seconds longer on answering the BDM version of the question, while they are almost 5 seconds faster for the binary responsibility question.[28] Hence, in the following regression analysis, we exclude those participants that misunderstood the elicitation procedure.

Table 4 reports results from regressions including the alternative measures of the responsibility and performance beliefs. Again, adding performance beliefs as controls is crucial since even same levels of responsibility attribution across group affiliations in the BDM can imply reverse discrimination. This would be the case if Germans had higher performance beliefs for other Germans than for refugees. The two-limit Tobit specification of column (1) includes *responsibility switchpoint* as dependent variable and the binary performance belief as control variable. We also control for the direction of shocks. The coefficient for *Refugee* is positive as before but now insignificant ($p = 0.393$), as opposed to in Table 1. Hence, also when controlling for beliefs and shock

[28]When designing the experiment, we decided against including control questions to ensure understanding of the BDM — as is often done for these complex elicitation procedures. We did not want to treat refugees and Germans differently because that by itself could have induced a treatment effect, and explaining the BDM in depth to the refugees would presumably have taken very long.

Table 4: Favorable responsibility attribution with continuous measures

Dependent variable	Favorable attribution		
	(1)	(2)	(3)
Refugee	0.216	0.119**	0.181
	(0.318)	(0.052)	(0.306)
Belief high	0.911***		
	(0.262)		
Switchpoint cutoff		0.113***	0.356***
		(0.011)	(0.090)
Neg shock	0.333	0.172***	0.339
	(0.226)	(0.047)	(0.258)
Constant	4.265***		2.590**
	(0.959)		(1.122)
Additional controls	Yes	Yes	Yes
Observations	140	142	131
Pseudo R^2	0.032	0.197	0.064

Notes: Column (1) and (3) report two-limit Tobit regressions on *responsibility switchpoint*. Column (1) includes the binary performance belief indicator *belief high*, whereas column (3) includes *performance switchpoint*. Column (2) reports average marginal effects of from a probit regression explaining *favorable attribution* with *performance switchpoint*. Subjects that clearly misunderstood the BDM elicitations are dropped. All columns include additional covariates from the questionnaire: age, semester, and number of experiments so far. Robust and clustered (on session level) standard errors in parentheses. Stars indicate significance on the levels: *** p<0.01, ** p<0.05, * p<0.1.

direction, we do not see a statistically significant positive effect of being matched with a refugee on responsibility attribution implied by the BDM elicitation. Using the binary responsibility measure and including non-binary performance beliefs in column (2), however, results in similar findings as in Table 1. The effect of *Refugee* is significantly positive. With both switchpoint variables instead of their binary counterparts in column (3), we again observe no significant reverse discrimination.

How can we explain the insignificant coefficients for the specifications using *responsibility switchpoint*? First, even when excluding inconsistent subjects, we still expect some misunderstanding in the BDM. Especially the BDM for responsibility attribution is rather difficult to grasp. This increases noise in the data and makes detecting the effect more difficult.

Second, indifference or only weak binary preferences are important. These weak inconsistencies, however, are still highly asymmetric. If only indifferent subjects were responsible for the different results of Table 1 and Table 4, a substantial fraction

of Germans matched with a refugee would have to be indifferent and attribute favorably in the binary elicitation, while those in *German* and indifferent would attribute unfavorably. This still is a clear form of reverse discrimination — it would only be less costly than if it was not driven by indifference. Similarly, other types of inconsistencies and choice reversals that we cannot categorize could drive the difference in our findings. We do have some evidence for this type of strong asymmetry in inconsistencies for the responsibility beliefs. Of the twelve participants being strictly inconsistent (light grey in upper panel of Table 3), five are subjects in *German* and all of these switch from unfavorable binary attribution to favorable switchpoint attribution. In stark contrast to that, of the seven strictly inconsistent Germans in *Refugee*, five switch from favorable binary attribution to unfavorable probabilistic attribution. Despite the very low number of observations, this is a significant difference ($p = 0.028$, Fisher's exact test, two-sided). The same is true for weak inconsistencies. For this purpose, we define those with a switchpoint of 5 in row (1) of Table 3 and a switchpoint of 6 in row (2) as being weakly inconsistent. In *German*, 12 out of 19 inconsistent subjects change from unfavorable binary to favorable switchpoint attribution, while only 9 out of 28 do so in *Refugee*. This difference is again significant ($p = 0.043$, Fisher's exact test, two-sided).

Third, with the BDM it might be more vague what the "right" thing to do is. If reverse discrimination is driven by self-image and identity concerns, the BDM elicitation procedure might well not make the identity prescriptions as clear as the binary elicitation. For the binary responsibility attribution it is obvious what the subjects should do if they do not want to blame someone. With probabilities this is less clear.

In summary, we get directionally very similar results with the non-binary belief elicitations. However, these results are weaker. Increased noise, indifference, systematic inconsistencies, and possibly increased opagueness of the normative prescription can help explaining this difference. While this provides some additional insights into individual decision making, it does not change our main message: We observe strongly asymmetric behavior leading to reverse discrimination and more

favorable treatment of refugees.

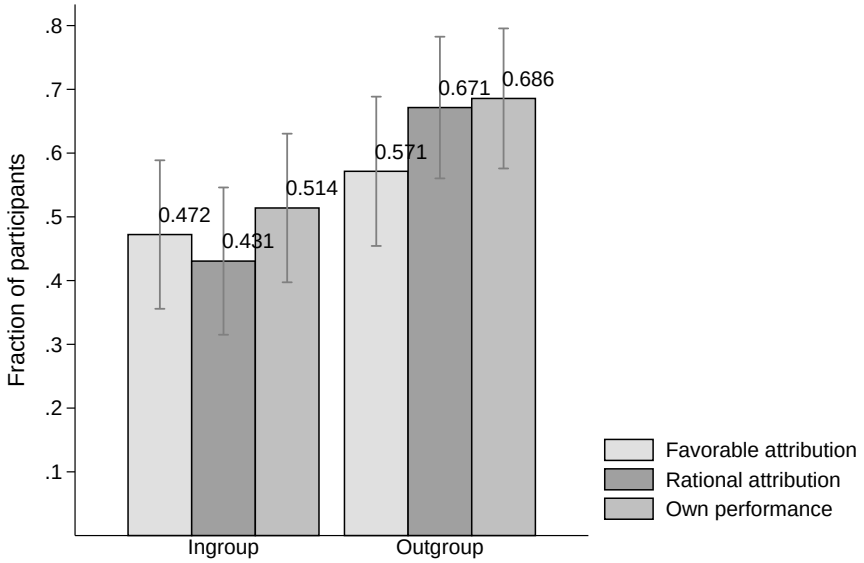
Result 4: *The evidence for reverse discrimination is weaker when considering non-binary beliefs. The asymmetry in behavior explaining this difference, however, again points to strongly group-specific patterns.*

4 The *KleeKandinsky Experiment*

In an additional experiment, we only invited participants from the regular subject pool and applied a minimal group paradigm to analyze whether our result of reverse discrimination is a general result for in- and outgroups or whether it stems from our specific groups in the *Refugee Experiment*. Since groups were formed based on preferences for paintings of the artists Klee and Kandinsky, henceforth we call this experiment *KleeKandinsky Experiment* (and our main experiment *Refugee Experiment*). With a total of 142 subjects, we ran six sessions in August 2016. Subjects earned €13.85 on average, including a €6 fixed payment for showing up on time. Each subject participated in one session only.

Procedures differed only in dimensions explicitly catered to refugees mentioned in Section 2. Hence, there was no gender restriction for participation, no Arabic announcements were made, participants only drew seat numbers from one bag, and group affiliation was communicated via group names (Klee or Kandinsky) instead of first names. Moreover, every subject is matched with only one other subject. Subjects in the *Ingroup* treatment ($n = 72$) are matched with a subject of the same group, while we match subjects of different groups with each other in the *Outgroup* treatment ($n = 70$).

We employ a modified version of the minimal group paradigm used by Chen and Li (2009). Subjects evaluate paintings of the artists Paul Klee and Wassily Kandinsky. Five pairs of paintings containing each a painting of Klee and Kandinsky are shown. For each pair and without knowing the artist of the paintings, participants have to decide which of the two paintings they prefer. Based on a median split in artist preferences, subjects are assigned to the Klee or Kandinsky group. This assignment procedure takes place at the very beginning of the experiment.



Notes: The figure shows *favorable attribution*, *rational attribution*, and *own performance* for the *KleeKandinsky Experiment*. Error bars indicate 95% confidence intervals.

Figure 6: *Favorable attribution*, *rational attribution*, and *own performance* in the *KleeKandinsky Experiment*

Contrary to the results of the *Refugee Experiment*, responsibility attribution is not affected by group affiliation of the matched partner in the *KleeKandinsky Experiment*. Figure 6 shows that attribution is more favorable in the *Outgroup* treatment (light gray bars), however, this can be explained by beliefs about performance. If anything, given rational attribution (dark gray bars), subjects in *Outgroup* should attribute responsibility even more favorably and subjects in *Ingroup* even less favorably. As can be seen from the intermediate gray bars at the very right, the difference in performance beliefs can be explained by differences in individual performances.[29]

Table 5 shows the same regression analysis as Table 1 does for the *Refugee Experiment*. As we already observed in Figure 6, in the baseline regression in column (1), it seems as if there is some form of reverse discrimination. This positive effect of being matched with an outgroup member is not robust to controlling for beliefs. The effect of group affiliation becomes a rather precise zero when we control for performance beliefs (see column (2)). In column (3), we include a dummy

[29]Even though individual performances should be orthogonal to treatment assignment, we still see pronounced differences. Participants in *Outgroup* solve 4.06 puzzles on average, while participants in *Ingroup* only solve 3.36 puzzles on average. This difference is significant ($p < 0.01$, Mann-Whitney *U*-test, two-sided). Table C.3 in the Appendix reveals that the sample is balanced otherwise. There are no differences with respect to age, number of semester, and number of experiments so far.

Table 5: Favorable responsibility attribution (*KleeKandinsky Experiment*)

Dependent Variable	Favorable attribution			
	(1)	(2)	(3)	(4)
Outgroup	0.099**	–0.006	0.023	0.010
	(0.038)	(0.057)	(0.061)	(0.056)
Belief high		0.392***	0.336***	0.345***
		(0.079)	(0.085)	(0.077)
Neg shock			0.258***	0.248***
			(0.057)	(0.056)
Additional controls	No	No	No	Yes
Observations	142	142	142	142
Pseudo R^2	0.007	0.141	0.206	0.224

Notes: Probit regressions on binary variable *favorable attribution* reporting average marginal effects. Column (4) includes additional covariates from the questionnaire: age, semester, and number of experiments so far. Robust and clustered (on session level) standard errors in parentheses. Stars indicate significance on the levels: *** p<0.01, ** p<0.05, * p<0.1.

for the direction of the shock. As in the *Refugee Experiment*, we find that subjects assign responsibility more favorably after negative shocks. Since shocks were evenly distributed across group affiliation in the *KleeKandinsky Experiment*,[30] we did not expect to observe an effect on the *Outgroup* coefficient. This is confirmed by column (3). Adding more controls in column (4) does not alter the results. Also note that effect sizes of *belief high* and *neg shock* are quite similar to the ones from the *Refugee Experiment*. Overall, this demonstrates that our finding of reverse discrimination is a result of our natural group assignment in the *Refugee Experiment* and not a general result in our experimental design.

Result 5: *There is no evidence for reverse discrimination with artificially assigned groups.*

5 Discussion

In this section, we discuss several explanations for why we find reverse discrimination in our setting. As we can rule out statistical discrimination, taste-based discrimination is a first natural candidate to look at. Subjects are willing to pay a price to attribute responsibility favorably towards refugees. In our context, taste-based discrimination

[30] 57% of subjects in *Outgroup* and 51% in *Ingroup* receive a positive income shock.

would imply that this is the case because they have some sort of preference for this group. This explanation seems, however, unlikely. First, participants matched with refugees do not affect refugees' payments by attribution behavior. Hence, outcome based tastes cannot play a role for choices. Second, the same holds for tastes for interaction. Participants never interact with their matched partner, and responsibility attribution choices do not affect the degree of interaction. Third, the results of the IAT reveal that Germans on average have negative implicit associations towards Arabic names. Lastly, taste-based explanations also stand in stark contrast to the literature on ingroup favoritism.[31]

The finding of favoring refugees might also be caused by the desire to be seen as a good person by others. Social image concerns have been shown to be an important motivation for decisions in various settings where behavior is publicly observable (e.g., Andreoni and Bernheim, 2009; Ariely et al., 2009; Lacetera and Macis, 2010). In our setting, however, subjects take their decisions completely anonymously, which is common knowledge to our subjects.[32] Similarly, our experimental results could be affected by experimenter demand effects (EDE), that is, in our case, by norm conformity pressure. While we cannot completely rule out such effects, some considerations render an interpretation of our results predominately based on this pressure unlikely. Participants could indeed perceive favorable attribution towards refugees as the appropriate behavior in the eyes of the experimenter. However, in our between-subjects design, EDE should have also affected behavior of our subjects in *German* and in the *KleeKandinsky Experiment*. This applies, in particular, to the *KleeKandinsky Experiment*. The artificiality of the minimal group paradigm (as opposed to a more natural identification based on first names) should, if anything, make EDE even more likely (as implied by Zizzo, 2010). In these other treatments though, beliefs about performance do not differ from favorable attribution. That is, behavior is in line with rational responsibility attribution leaving the *Refugee* treatment as the

[31] See, e.g., the literature review by Hewstone et al. (2002).

[32] At the beginning of the experiment, we guarantee our subjects that all of their decisions will be analyzed anonymously. The experimenter is not present in the laboratory while decisions are taken. In addition, it is not possible to infer decisions directly from the level of payoffs (which is observed by the research assistant privately handing out the earned money).

only biased sample.[33] Importantly, both social image concerns and norm conformity pressure — if they occurred in our experiment — are likely to more strongly occur in non-anonymous decision environments. Compared to actual behavior in the field, our results would then provide a lower bound.

In addition to being motivated by appearing as a good person in front of others, one could be motivated by appearing as a good person in front of oneself. Keeping up a certain identity, a person's self-view, oftentimes conflicts with profit maximizing behavior and explains departures thereof in different economic spheres (e.g., Akerlof and Kranton, 2000; Mazar et al., 2008). This can also lead to deliberately distorted beliefs, i.e., motivated beliefs (e.g., Di Tella et al., 2015; Gneezy et al., 2016; Grossman and Van Der Weele, 2017). Agents with such motivated beliefs have a positive willingness to pay for keeping up a specific self-image. We find that our subjects make choices that are in line with behaving "politically correct". Especially with regard to our student subject pool, it seems to be plausible that being open and tolerant towards minorities is part of our subjects' identity. In order to keep up a positive self-view, they seem to be reluctant to blame refugees. There is some evidence from psychology supporting such reasoning. Dutton (1973) finds that middle-class Canadian whites donate more when the solicitor is of black or Indian ethnicity as compared to when the solicitor is white. With donors perceiving black people and Indians to be targets of discrimination, the author interprets the results as supportive evidence for a specific type of revealed reverse discrimination. In simple interactions, minority groups will be treated better than other ingroup members. In addition, Byrd et al. (2015) show that liberal and moderate whites favor black over white politicians in an artificial setting. Participants read political speeches and saw a picture of either a black or a white person who was supposed to have given the speech. Among other outcome variables, more participants indicated that they would vote for a black politician. The evidence of these studies suggests that actively avoiding explicit discrimination might be part of the identity of politically liberal and moderate middle-class people to which

[33] At the end of the experiment, we further ask for non-incentivized verbal explanations for behavior. We do not have a single statement that could be related to EDE.

the majority of our subjects should belong to. This explanation is also in line with the stronger results for the binary responsibility beliefs compared to the finer-graded probability beliefs. In the former elicitation, it is absolutely clear what the "good" or "bad" thing to do is. Hence, our subjects try to avoid taking the bad action towards the refugees.[34] In contrast, "good" and "bad" is not as clearly defined for the latter elicitation procedure. We therefore argue that motivated belief formation is the most plausible explanation for our main result.

6 Conclusion

We experimentally study responsibility attribution for negative and positive income shocks. In particular, we ask whether there is asymmetric attribution of responsibility, depending on whether a German participant is matched with another German or a refugee. In our setting, there is imperfect information regarding the source of the shock. It can either be due to a random draw or due to the performance of the matched participant. This experimental paradigm is an abstract setting related to several environments in the field. Oftentimes, there is uncertainty with regard to what or who is responsible for a certain outcome. Group-specific behavior can thus strongly impact the lives of different societal groups. Prominent examples relate to labor market settings, where people that are discriminated against in responsibility attribution will be strongly disadvantaged. This might occur in the hiring process or at later stages in promotion, job assignment, or bonus decisions. Our study also relates on a more aggregate level to how developments and outcomes for the society as a whole might be related to groups of people. Recent examples are the strongly debated effects of refugees on crime, economic prospects of societies, and cultural developments. The negative shock of rising crime rates in some European countries might be indeed (in part) caused by the influx of refugees (as suggested by Donald Trump's quote at the

[34]We further assume that there is a clear difference in moral prescriptions between stating performance beliefs and responsibility beliefs. While it should be perceived a good (bad) thing to praise (blame) for responsibility, there should be no such moral connotation to stating mere performance beliefs. This is why we expect to observe distorted (discriminating) responsibility attribution and rather unbiased performance beliefs.

beginning of this paper) but could also be due to many other factors.

Surprisingly and contrary to the literature, which predominantly documents ingroup favoritism, we find no discrimination against refugees in responsibility attribution. Importantly, refugees are clearly not blamed for negative events but less often held responsible when a negative shock occurs. That is, we observe reverse discrimination. German participants generally attribute responsibility to refugees more favorably as compared to other Germans. We put forward an explanation based on identity concerns and motivated beliefs. Participants want to view themselves as non-xenophobic and tolerant and hence distort attribution as to not conflict with this identity. This belief distortion consequently leads to reverse discrimination. Comparing these results to an experiment with artificial group assignment, we show that our results are not a general result for in- and outgroups but rather depend on our specific sample. This lends support to the idea that the refugee sample indeed induces identity concerns. Furthermore, implicit associations of our German participants towards Arabic names are negative, while responsibility attribution is irrationally favorable on average. This suggests that favoring refugees is a conscious choice in our experiment. Moreover, we find that subjects with more positive associations towards Arabic names attribute responsibility more favorably to them. Implicit associations — which are correlated with important field behavior such as hiring decisions — thus predict responsibility attribution in a meaningful way.

The evidence for reverse discrimination towards refugees together with our results on potential mechanisms provide fruitful avenues for future research. First, while we find strong evidence in the domain of responsibility attribution, our study cannot draw conclusions about whether our finding for the natural outgroup of refugees translates into other domains of discrimination such as trust or social preferences. Second, our sample of university students (in Munich) is not representative for the population (of Germany). This has implications for the generalizability of our results. Similar studies with more right-wing and less liberal subpopulations might yield different results. Hence, testing our findings with different subject pools can yield additional insights — especially with regards to the effect of

identity concerns. Future research could also exogenously vary identity concerns by priming certain aspects of subjects' identities. This could help to establish a causal link between these concerns and discrimination behavior. Lastly, the difference between our findings in the binary versus the probability-scale responsibility attribution highlight a potentially mediating effect of moral prescriptions. Using a range of choice environments that differ in the strength of behavioral prescriptions could test this relationship.

References

Ai, C. and Norton, E. C. (2003). "Interaction terms in logit and probit models." *Economics Letters*, 80(1), 123–129.

Akerlof, G. A. and Kranton, R. E. (2000). "Economics and identity." *Quarterly Journal of Economics*, 115(3), 715–753.

Andreoni, J. and Bernheim, B. D. (2009). "Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects." *Econometrica*, 77(5), 1607–1636.

Ariely, D., Bracha, A. and Meier, S. (2009). "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially." *American Economic Review*, 99(1), 544–55.

Bartling, B. and Fischbacher, U. (2011). "Shifting the blame: On delegation and responsibility." *Review of Economic Studies*, 79(1), 67–87.

Bartling, B., Fischbacher, U. and Schudy, S. (2015). "Pivotality and responsibility attribution in sequential voting." *Journal of Public Economics*, 128, 133–139.

Becker, G. M., DeGroot, M. H. and Marschak, J. (1964). "Measuring utility by a single-response sequential method." *Behavioral Science*, 9(3), 226–232.

Byrd, D. T., Hall, D. L., Roberts, N. A. and Soto, J. A. (2015). "Do politically non-conservative whites "bend over backwards" to show preferences for black politicians?" *Race and Social Problems*, 7(3), 227–241.

Carneiro, P., Heckman, J. J. and Masterov, D. V. (2005). "Labor market discrimination and racial differences in premarket factors." *Journal of Law and Economics*, 48(1), 1–39.

Chen, Y. and Li, S. X. (2009). "Group identity and social preferences." *American Economic Review*, 99(1), 431–457.

Di Tella, R., Perez-Truglia, R., Babino, A. and Sigman, M. (2015). "Conveniently upset: Avoiding altruism by distorting beliefs about others' altruism." *American Economic Review*, 105(11), 3416–3442.

Dutton, D. G. (1973). "Reverse discrimination: The relationship of amount of perceived discrimination toward a minority group on the behaviour of majority group members." *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 5(1), 34–45.

Fershtman, C. and Gneezy, U. (2001). "Discrimination in a segmented society: An experimental approach." *Quarterly Journal of Economics*, 116(1), 351–377.

Fischbacher, U. (2007). "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics*, 10(2), 171–178.

Gneezy, U., Saccardo, S., Serra-Garcia, M. and van Veldhuizen, R. (2016). "Motivated self-deception, identity and unethical behavior." *Working Paper*, mimeo.

Goldin, C. and Rouse, C. (2000). "Orchestrating impartiality: The impact of "blind" auditions on female musicians." *American Economic Review*, 90(4), 715–741.

Greene, W. (2010). "Testing hypotheses about interaction terms in nonlinear models." *Economics Letters*, 107(2), 291–296.

Greenwald, A. G., Uhlmann, E. L., Poehlman, T. A. and Banaji, M. R. (2009). "Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity." *Journal of Personality and Social Psychology*, 97(1), 17–41.

Greiner, B. (2015). "Subject pool recruitment procedures: organizing experiments with ORSEE." *Journal of the Economic Science Association*, 1(1), 114–125.

Grossman, Z. and Van Der Weele, J. J. (2017). "Self-image and willful ignorance in social decisions." *Journal of the European Economic Association*, 15(1), 173–217.

Heckman, J. J. (1998). "Detecting discrimination." *Journal of Economic Perspectives*, 12(2), 101–116.

Hewstone, M. (1990). "The 'ultimate attribution error'? A review of the literature on intergroup causal attribution." *European Journal of Social Psychology*, 20(4), 311–335.

Hewstone, M., Rubin, M. and Willis, H. (2002). "Intergroup bias." *Annual Review of Psychology*, 53(1), 575–604.

Jones, E. E. and Harris, V. A. (1967). "The attribution of attitudes." *Journal of Experimental Social Psychology*, 3(1), 1–24.

Krieger, N. (2014). "Discrimination and health inequities." *International Journal of Health Services*, 44(4), 643–710.

Lacetera, N. and Macis, M. (2010). "Social image concerns and prosocial behavior: Field evidence from a nonlinear incentive scheme." *Journal of Economic Behavior & Organization*, 76(2), 225–237.

Lane, T. (2016). "Discrimination in the laboratory: A meta-analysis of economics experiments." *European Economic Review*, 90, 375–402.

Lang, K. and Manove, M. (2011). "Education and labor market discrimination." *American Economic Review*, 101(4), 1467–1496.

Mazar, N., Amir, O. and Ariely, D. (2008). "The dishonesty of honest people: A theory of self-concept maintenance." *Journal of Marketing Research*, 45(6), 633–644.

Ockenfels, A. and Werner, P. (2014). "Beliefs and ingroup favoritism." *Journal of Economic Behavior & Organization*, 108, 453–462.

Pettigrew, T. F. (1979). "The ultimate attribution error: Extending Allport's cognitive analysis of prejudice." *Personality and Social Psychology Bulletin*, 5(4), 461–476.

Ross, L. (1977). "The intuitive psychologist and his shortcomings: Distortions in the attribution process." *Advances in Experimental Social Psychology*, 10, 173–220.

Shapiro, T., Meschede, T. and Osoro, S. (2013). "The roots of the widening racial wealth gap: Explaining the black-white economic divide." *Research and Policy Brief*.

Zizzo, D. J. (2010). "Experimenter demand effects in economic experiments." *Experimental Economics*, 13(1), 75–98.

Appendix

A Refugee Recruiting Details

Refugees were recruited by distributing the leaflet shown in Figure A.1. The actual first names of the refugees taking part in the experiment and which were visible to the matched partner were: Abdo, Abduh, Abdullah (2x), Adnan, Ahmad (3x), Alaa, Ali, Alkhder, Almhklf, Amjad, Anas, Bshr, Firas, Ghassan, Ghiath, Giwan, Hafez, Hasan, Khaled (2x), Louay, Mazen (2x), Mohamad, Mohamd, Mohammad, Mohammed (3x), Mounir, Nizar, Obaida, Odai, Omar, Sabri, Saleem, Schindar, Wissam, Yazan, Youssef.

Figure A.1: Leaflet for recruiting refugees (translated from Arabic)

The names of the German participants were: Aleksandar, Alex, Alexander (3x), Aljoscha, Andi, Andreas (2x), Axel, Ben, Benedikt, Benjamin, Benno, Bernhard, Caspar, Chris, Christian (3x), Christoph, Christopher, Daniel (4x), David (4x), Dominic, Dominik (2x), Eric, Fabian (7x), Felix (3x), Fiete, Florian (2x), Franz, Franziskus, Fridtjof, Gregor, Ion, Jan, Jan Fedor, Jens, Joel, Johannes (4x), Jonas (3x), Jonathan

(2x), Josaphat, Julian (3x), Kevin, Konstantin (2x), Korbinian (2x), Laurian, Lennart, Leon, Leonard, Lion, Louis, Lukas (2x), Manuel, Marcus (3x), Marian, Marius (4x), Markus (3x), Martin (2x), Matthias (5x), Maurus, Max (5x), Maximilian (3x), Michael (4x), Moritz, Niclas, Niklas, Niko, Oswald, Pascal, Patrick, Paul, Philipp (4x), Raffael, Richie, Roman, Sebastian (3x), Simon, Stefan (3x), Steffen, Stephan (2x), Thomas (3x), Tilman, Tim, Timo, Tobi, Tobias (3x), Tom, Valentin, Vincent.

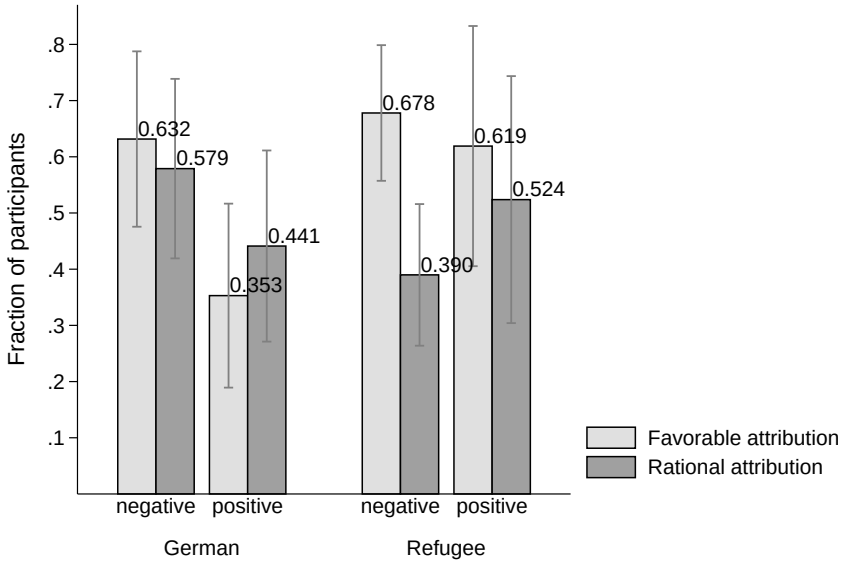
B Puzzle Motives

The selected motives for the puzzles are pictures of a range of colors, a bird, a beach, a lamb, a tree in a desert, a sunset over the ocean, a water drop, and a box of bananas. They are displayed in Figure B.1.

Figure B.1: Puzzle motives for real effort task

C Supplementary Results

C.1 Responsibility Attribution by Shock



Notes: The figure shows *favorable attribution* and *rational attribution* for both treatments divided by shock direction. Error bars indicate 95% confidence intervals.

Figure C.1: *Favorable attribution* and *rational attribution* by shock direction

Figure C.1 shows actual attribution behavior and counterfactual rational attribution based on performance beliefs for both group affiliations by shock direction. Even though, at first glance, it looks as if behavior in *Refugee* after a negative shock drives reverse discrimination, comparing behavior across the two group affiliation shows that the difference in difference is rather similar for both shocks. After a negative shock, participants in *Refugees* deviate by 0.288 from rational attribution, while those in *German* attribute responsibility more favorably by 0.053. This is a difference in difference of 0.235. After a positive shock, the deviation for participants in *Refugees* is 0.095 and -0.088 in *German*. Hence, the difference in difference sums up to 0.183, and is therefore close to 0.235 after a negative shock.

C.2 Balance Table *Cond* vs. *Uncond*

Table C.1: Balance table *Refugee Experiment* (*Cond* vs. *Uncond*)

	Cond (1)	*Uncond* (2)	(1) vs. (2) p-value
Own performance	0.368	0.579	0.009
Age	22.474	23.303	0.160
Semester	4.224	4.553	0.534
Number of experiments so far	5.461	8.250	0.021

Notes: Own performance indicates whether a subject solved four or more puzzles.

C.3 Regression Analysis Controlling for Own Performance

Table C.2 reports results from regressions equivalent to our main regressions in Table 1 (Section 3.1) only using the number of correctly solved puzzles as control variable instead of performance beliefs directly.

Table C.2: Favorable responsibility attribution (controlling for own performance)

Dependent variable	Favorable attribution			
	(1)	(2)	(3)	(4)
Refugee	0.160***	0.181***	0.144***	0.139***
	(0.056)	(0.055)	(0.047)	(0.044)
# correct puzzles		0.089***	0.086***	0.091***
		(0.022)	(0.023)	(0.022)
Neg shock			0.159**	0.148**
			(0.063)	(0.064)
Additional controls	No	No	No	Yes
Observations	152	152	152	152
Pseudo R^2	0.020	0.062	0.081	0.090

Notes: Probit regressions on binary variable *favorable attribution* reporting average marginal effects. Column (4) includes additional covariates from the questionnaire: age, semester, and number of experiments so far (all insignificant). Robust and clustered (on session level) standard errors in parentheses. Stars indicate significance on the levels: *** p<0.01, ** p<0.05, * p<0.1.

C.4 Balance Table for the *KleeKandinsky Experiment*

Table C.3: Balance table *KleeKandinsky Experiment*

	Ingroup (1)	*Outgroup* (2)	(1) vs. (2) p-value
Own performance	0.514	0.686	0.037
Age	24.875	24.729	0.842
Semester	5.736	5.129	0.220
Number of experiments so far	10.542	11.700	0.401

Notes: Own performance indicates whether a subject solved four or more puzzles.

D Interaction Effect of IAT Score and Being Matched with a Refugee

For estimating the interaction effect between having a negative IAT score and our treatment, we compute predictive values for *favorable attribution* by using probit regression estimates from model (3) used in Table 2 for the following four groups:

- Subjects in *Refugee* with a negative IAT score:
  $\widehat{P(Y = 1|Refugee = 1, IAT < 0, X)} = 0.5862$
- Subjects in *Refugee* with a positive IAT score:
  $\widehat{(Y = 1|Refugee = 1, IAT > 0, X)} = 0.8375$
- Subjects in *German* with a negative IAT score:
  $\widehat{P(Y = 1|Refugee = 0, IAT < 0, X)} = 0.5295$
- Subjects in *German* with a positive IAT score:
  $\widehat{P(Y = 1|Refugee = 0, IAT > 0, X)} = 0.4189$

This leaves us with a difference in differences of –0.3619 ([0.5862 – 0.8375] – [0.5295 – 0.4189]). Thus, the effect of having a negative IAT score on *favorable attribution* is 36.19 percentage points lower in *Refugee* than in *German*.

E Instructions

The following passages are the instructions for *Cond* translated from German. Text in italics refers to instructions read out aloud by the experimenter (alternating one of the two authors), which were repeated in Arabic. Text in brackets indicates self-explaining comments. Text in normal letters refers to instruction that the subjects read on screen (either in German or Arabic).

[upon arrival at the laboratory]
Hello everybody. We provide refugees with the possibility to take part in a series of experiments. This is why there are refugees among the participants today. In order to assign you to the seat with the correct language [experimenter points at the two bags labeled with "German" or "Arabic"] *Arabic-speaking participants draw a card with a seat number from the bag with the label Arabic and German-speaking participants a card from the bag with the label German.*

[in the laboratory after seating took place]
Welcome to MELESSA. Thank you very much for showing up to this experiment on time. My name is Felix Klimm/Stefan Grimm, and I will conduct this experiment today.

Please do not talk to other participants during the experiment.

For the sake of simplicity, you find the instructions on your screen. The instructions are the same for all participants. Please follow the instructions. If you have any questions, please raise your hand or press the red button on your keyboard. We will then come to you and answer your question in private.

[first screen]
General Procedures I
This experiment is meant to study economic decision making. It will last about 1 hour. You can earn money during the experiment. This money will be paid to you in private after the experiment. You will make decisions in this study. These decisions will affect your payment. In addition, your payment might depend on other participant's

decisions as well as on chance. Further rules will be explained to you right before each decision. Hence, today's payment is the sum of money earned with your decisions plus €6 for showing up on time.

[new screen]

General Procedures II

The experiment consists of 2 parts. You will see the instructions for each part right before the respective part starts. Data from this experiment will be analyzed anonymously. At the end of the experiment, you will have to sign a receipt. This is only for accounting purposes.

[new screen]

Part 1

In part 1 of the experiment, you need to perform a task. You receive €3 for performing this task. Your task is to correctly solve as many puzzles as possible. This task is suited for everybody as puzzles are well known in most parts of the world. For this purpose, there are 8 puzzles next to your keyboard. You are allowed to start as soon as we tell you to do so. After 10 minutes, you need to stop, and we will count the number of correct puzzles. There will be a clock on your screen displaying the remaining time. Click on OK if you understand the procedure. Please still wait with solving a puzzle until we tell you to start.

[Subjects perform real effort and the experimenter and student research assistants checks the number of correctly solved puzzles.]

[new screen]

Part 2

You are now matched with another participant. Please enter your first name for this purpose. Thereafter, the first name of your matched participant will be shown to you. Your matched participant will see your first name.

Your first name: ≪own name≫

[new screen]

Your matched participant is: ≪name partner≫

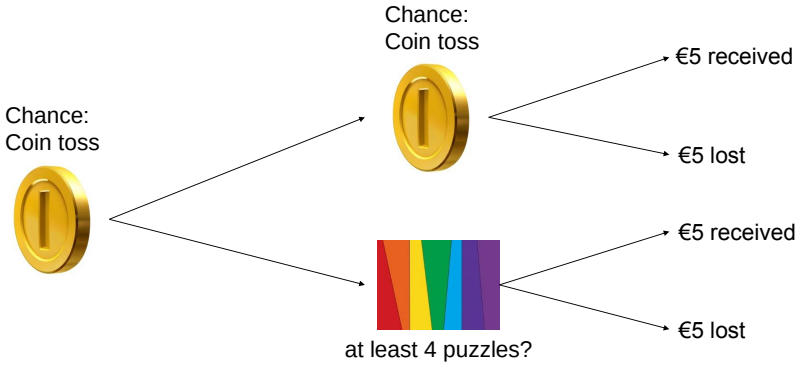
[new screen]

Your payoff might depend on your matched participant's decisions. Reminder: Your matched participant is ≪name partner≫. In the following, you can receive additional €5 or lose €5. Whether you are receiving or losing €5 depends on chance or the other participant. First, the computer will determine via a virtual coin flip whether chance or the other participant is responsible for your payment. Both cases are equally likely (50/50). Hence, there are 2 possibilities:

1. If chance is responsible, you will receive €5 with 50% probability. Hence, a coin will be flipped again.

2. If ≪name partner≫ is responsible, the number of puzzles that ≪name partner≫ solved correctly in part 1 will determine whether you receive or lose €5. If ≪name partner≫ solved at least 4 puzzles, you will receive €5. If ≪name partner≫ solved fewer than 4 puzzles, you will lose €5.

The graph below illustrates the procedure.



[new screen]

You will know about your payment in a second. However, you will not know whether chance or ≪name partner≫ is responsible for this payment.

Please answer four test questions in order to be sure that you understand the procedure.

[new screen]

1. If ≪name partner≫ solved at least 4 puzzles, will you receive €5 in any case?

2. If ≪name partner≫ solved 3 or fewer puzzles and chance was selected to be responsible for your payment, how likely is it that you will receive €5?

3. If chance was selected to be relevant for your payment, does your payment depend on the number of correctly solved puzzles by ≪name partner≫ in this case?

4. How much lower will your payment be if you lose €5 compared to the case in which you receive €5?

[new screen]

You have answered all the questions correctly. On the next screen you will see whether you receive or lose €5.

[new screen]

Your income:

Reminder: The computer randomly determined whether chance or ≪name partner≫ is relevant for your payment. According to these rules:

You receive/lose €5.

[new screen]

We now ask you to answer 4 questions. One of the questions will be randomly selected at the end of the experiment. You will then receive payment according to your answer to this question.

[new screen]

Question 1

Do you believe that chance or ≪name partner≫ was responsible for your payment?

If your answer is correct and this questions will be selected to be payoff relevant, you receive €5.

[new screen]

Question 2

You will now make a sequence of decisions. Each of the decisions contains 2 options — A and B. Both options give you once more the chance to receive another €5.

One of the 9 rows will be randomly chosen for payment if question 2 will be payoff relevant.

If you choose option A in one of the 9 rows, you will receive €5 if ≪name partner / chance≫ [name of partner or chance displayed depending on the answer to Question 1 — name of the partner displayed if subject indicated that the partner is responsible] was responsible for your payment.

If you choose option B, you will receive €5 with a certain probability. This probability varies from 10 to 90 percent and is shown to you next to every decision.

If question 2 is payoff relevant, one of your 9 decisions will be implemented. The computer will randomly select which decision will be implemented in this case.

Please consider now from which probability on (which row) you want to choose option B. If you took your decision, click on OK.

Option A You receive €5 if ≪name partner / chance≫ [here, again, name of partner or chance displayed depending on the answer to Question 1] was responsible for your payment.

Option B You receive €5 with a probability of 10% ... 90%.

[new screen]

Question 3

Do you believe that ≪name partner≫ solved at least 4 puzzles? Hence, did he solve 4, 5, 6, 7, or 8 puzzles?

If your answer is correct and this questions will be selected to be payoff relevant, you receive additional €5.

[new screen]

Question 4

In question 4 — like in question 2 — you will make a sequence of decisions. Each of the decisions contains 2 options — A and B. Both options give you the chance to receive another €5.

One of the 9 rows will be randomly chosen for payment if question 4 will be payoff relevant.

If you choose option A in one of the 9 rows, you will receive €5 if ≪name partner≫ solved at least 4 puzzles.

If you choose option B, you will receive €5 with a certain probability. This probability varies from 10 to 90 percent and is shown to you next to every decision.

If question 4 is payoff relevant, one of your 9 decisions will be implemented. The computer will randomly select which decision will be implemented in this case.

Please consider now from which probability on (which row) you want to choose option B. If you took your decision, click on OK.

Option A You receive €5 if ≪name partner≫ solved at least 4 puzzles.

Option B You receive €5 with a probability of 10% ... 90%.