Queen Bee Immigrant:
The effects of status perceptions on immigration attitudes

Biljana Meiske*

January 5, 2025

Abstract

This work examines a seemingly counter-intuitive phenomenon observed in many Western democracies, whereby (parts of) the immigrant population supports anti-immigration policies. I propose that relative status deprivation, that is, the degree to which a given national/ethnic group is ranked low in the ethnic status hierarchy of the host country, has a negative impact on the attitudes of its members toward even lower-ranked groups. In an experiment, participants with an immigration background residing in Germany (N=1,159) receive either a positive or a negative evaluation of their national/ethnic in-group, while holding fixed evaluations of several outgroups, as evaluated by a group of native-majority (German) participants. Receiving a negative evaluation of their in-group leads participants to express more negative views of the refugees from the Middle East and to significantly decrease willingness to donate to an organization supporting refugees, while not altering their generosity in a general setting unrelated to immigration.

JEL Codes: C90, J15, J71

Keywords: Immigration attitudes, Discrimination, Status

Pre-Registration: AEARCTR-0008630

Ethics Clearance: Approved by Ethics Commission of the Department of Economics at Ludwig Maximilian University, Munich

Funding: Funding for this work was provided by the Max Planck Society.

*European University Institute biljana.meiske@eui.eu

I am grateful to Sule Alan, Davide Cantoni, Elias Dinas, Florian Englmaier, Vicky Fouka, Lata Gangadharan, Andrea Ichino, Andrea Martinangeli, Simeon Schudy, Raisa Sherif, Mirko Wiederholt, as well as participants of research seminars in economics and political science at the European University Institute, economics department of Ludwig Maximilian University and research seminar at the Max Planck Institute for Tax Law and Public Finance, for many helpful comments.

1 Introduction

In the 2017 federal election in Germany, Alternative für Deutschland (AfD), reached higher electoral support among the Russian-speaking Germans compared to the national average (Goerres et al. (2020)). The support among this community for a populist right-wing party that based the core of its platform on opposing immigration attracted much media attention as a curious phenomenon. Indeed, this is seemingly counter-intuitive – why would groups who themselves have a history of immigration, and are also largely perceived by natives as immigrants, support anti-immigration platforms? This is, however, by far not the sole example of opposition to new waves of immigration expressed by immigrants. Consider, for instance, the 2024 U.S. presidential election and the numerous reports of the surprising success of Donald Trump among the Latino community (Lange et al. (2024)). While other issues might have contributed to Trump's appeal within this electorate, earlier evidence documented explicit support for restrictive immigration policies among considerable segments of this group (Lopez and Taylor (2010)). Similarly, the "Leave" option in the Brexit referendum garnered significant support among the British South Asian population, with accounts often pointing to concerns over new waves of immigration as explaining this preference (Shackle (2017), Dunin-Wasowicz (2017), Dunin-Wasowicz (2016)). In Switzerland, voters with immigrant backgrounds supported the referendum "against mass immigration" as much as the native majority did (Strijbis and Polavieja (2018)). Prominent politicians of immigrant descent, such as Suella Braverman, Nicky Haley, and João Varela, who built careers as outspoken opponents of immigration, further exemplify the phenomenon[1].

While a vast literature studies the attitudes of the majority population toward immigration (for a survey of this literature, see, e.g., Alesina and Tabellini (2022) and Hainmueller and Hopkins (2014)), much less attention is paid to the immigration attitudes among immigrants and their determinants. This is remarkable, given that immigrants constitute an increasing share of the (voting) population in virtually all Western democracies and considering the importance of immigration attitudes in

[1]Further examples are provided by, e.g., Pettersson et al. (2016) in Sweeden, or Meeusen et al. (2019) in case of Belgium.

shaping broader political preferences. This paper studies immigration attitudes among immigrants and examines the role of status enjoyed by a given immigrant group in the host society as a driver of their attitudes toward other immigrant minorities. I propose that relative status deprivation, that is, the degree to which an immigrant group is ranked low in the ethnic status hierarchy of the host country, has a negative impact on its members' attitudes toward groups ranked even lower[2]. In other words, I suggest that immigrant groups encountering an unfavorable reception in the host society respond to this experience by kicking down the ladder. In a frequently encountered situation, where the new-coming immigrants are perceived more critically and occupy a lower status position compared to the established ones, this amounts to status assigned to the established immigrants, shaping their attitudes towards new waves of immigration.

To investigate this idea, I run a preregistered survey-experiment with a sample of immigrant population in Germany. I experimentally vary the status perception of participants' immigrant in-group and examine the effect of this manipulation on their attitudes toward an immigrant out-group ranked even lower in the status hierarchy. To do this, I first conduct a pre-study where a sample of native German participants are asked to evaluate different immigrant groups (structured along the geographical region of their origin) and their contribution to life in Germany. The collected evaluations serve to create the treatments in the main experiment.

In the main experiment, participants with an immigration background ($N = 1,159$) are randomly allocated to be presented an evaluation collected in the pre-study, which, depending on the treatment, evaluates *their own immigrant in-group* as contributing either "rather positively" or "rather negatively" to the "*socio-economic and cultural life in Germany*". Participants are informed about the pre-study and know that the evaluations stem from native Germans (with no immigration background). While presenting participants with a single evaluation - the one regarding their own immigrant in-group - should be sufficient to administer the shock in perceived group status, this could entail an ambiguity in interpreting presented information. Participants could

[2]Whereas status as a concept has been used to designate group's ranking along different valued dimensions (e.g., socio-economic status), in this work, I rely on the definition of status as a ranking in terms of social esteem, honor, and respect accorded to them, distinct from wealth or power (Ridgeway (2019), Weber (1968)). Therefore, the status position of an immigrant group here refers to its position in the ethnic hierarchy of the host society, i.e., how socially desirable the group is perceived to be relative to other ethnic groups (importantly, including the native majority).

understand the presented information as a signal of the broader sentiment of the native majority toward all immigrants rather than being specific to their own group and believe that it is a matter of coincidence that exactly their in-group is mentioned. To address this concern and target the evaluation more precisely to the participant's in-group, another two evaluations of two other immigrant out-groups (one positive and one negative) are presented in the same way in both treatments. Therefore, participants in both treatments observe three evaluations, two of which remain always the same, and the only between-treatment variation comes from participants' own immigrant in-group being evaluated either positively or negatively.

To investigate the effect of manipulating the status perception of one's own in-group on their attitudes toward an out-group enjoying lower status, I use the donation game to measure participants' support for the refugees from the Middle East and Africa. During the so-called "refugee crisis" of 2015 and the following years, Germany received large numbers of asylum seekers fleeing conflicts in Syria, Iraq, and Afghanistan, which sparked a prolonged heated public debate, rendering the refugees a salient immigrant group[3]. Furthermore, results of multiple studies of attitudes toward minorities in Germany and other Western European countries demonstrate that immigrants from the Middle East and Africa, are perceived more critically compared to immigrants of (Eastern or Southern) European descent and can credibly be assumed to occupy a lower status position in the ethnic hierarchy (see e.g. Froehlich and Schulte (2019), Meidert and Rapp (2019), Sahgal et al. (2018); this finding was additionally confirmed in the present study). Participants' support for the refugees is captured by their willingness to forgo some part of their experimental earnings in order to secure a donation to the United Nations High Commissioner for Refugees (UNHCR). I additionally collect several attitudinal measures of participants' positions toward refugees.

The findings support the hypothesis that status deprivation has a negative impact on immigrants' attitudes toward refugees. Participants who received a negative evaluation of their own in-group on average pledged to donate 35 euros to the UNHCR (maximum

[3]The study was conducted before the beginning of the Russian invasion of Ukraine, so that the public discussion surrounding refugees effectively synonymized the term "refugees" with the asylum seekers from the Middle East and Africa. Indeed, according to the data of the Federal Office for Migration and Refugees of Germany, in 2020, more than 50% of asylum requests in Germany came from citizens of Syria, Iraq, or Afghanistan, and around 18% came from citizens of one of the countries in Africa.

possible amount being 100 euros). The average pledged donation among those who received a positive evaluation of their in-group was 11% higher (38.9 euros), a difference that is both sizeable[4] and statistically significant (one sided t-test for positive difference: $t = 1.95$, $p = 0.026$). I additionally show that, while the treatment variation does not affect participants' beliefs relating to topical concerns frequently discussed in the context of the immigration of refugees (e.g., their contribution to criminality, unemployment, or fiscal burden that they impose), exposure to negative prejudice leads to a decrease in willingness to allow more refugees into the country. These results are robust to controlling for various individual characteristics, including age, gender, education, household income, region of residence (in Germany), and region of origin of established immigrants.

My design allows me to rule out the mechanism whereby the treatment effect results from delivering a stark negative (positive) message to the participants, which might simply set them in a bad (good) mood, thereby affecting their prosociality more generally. To do this, I elicit participants' post-treatment generosity in a situation unrelated to immigration, captured by giving in a dictator game with an anonymous other participant. The results, show that, in difference to the donations benefiting the refugees, the dictator giving is not affected by the treatment variation. Additionally, a collected measure of participants' post-treatment mood shows that the treatment did not affect participants' mood.

Furthermore, I show that the treatment effect is not propagated through manipulation of the perceived economic threat stemming from the immigration of refugees. Recall that the evaluations describe participants' in-group as contributing either positively or negatively to the socio-economic and cultural life in Germany. It is thus possible that the statements might affect participants' perception of the socio-economic standing of their in-group (or merely remind them thereof) and thereby possibly reinforce (or weaken) the perceived threat of intensified competition in the labor market, or for welfare benefits, that might result from the immigration of refugees. To examine this possibility, I compare participants' answers to a question asking them if they believe that "refugees take away our jobs and social benefits" and

[4]For comparison, Dinas et al. (2021) measure the same outcome variable among German participants elicited in the identical way, and find (within control group) the difference between the center-left (SPD) voters and center-right (CDU/CSU) voters amounting to 12.2%.

show that the answers do not differ between the treatments. Moreover, taking income and education as proxies of participants' socio-economic standing, I show that the treatment effect is not stronger (and if anything is somewhat weaker) among immigrants with lower income and education, who can be expected to be particularly susceptible to reacting to the economic threat.

I additionally show that reciprocal preferences cannot explain the observed result. Specifically, I investigate the possibility that participants respond to encountered negative (positive) attitudes from the native majority by expressing a more negative (positive) attitude toward the refugees. Such behavior would be in accordance with the concept of upstream indirect reciprocity - a tendency of individuals to exhibit prosocial behavior toward someone because someone else has behaved prosocially toward them - an idea that has already found some empirical support (Alexander (1987), Nowak and Sigmund (2005)). I collect a measure of participants' indirect reciprocity preferences in an extended dictator game. I show that whereas participants with higher reciprocity preferences do react more to the negative (but not the positive) treatment relative to those lower in reciprocity, the effect of the treatment is not explained by accounting for this regularity.

My results are consistent with the predictions of the Social Identity Theory (SIT) (Tajfel et al. (1979)). SIT states that individuals in part define their own identities with regard to the social groups that they belong to (Tajfel (1969)). In order to facilitate a positive evaluation of oneself in the society, individuals are motivated to achieve a positive comparison of their group relative to the other ones (Turner (1975)). Therefore, being assigned to a low-status group presents an identity threat for its members, who are in turn predicted to engage in defensive strategies in order to cope with it. One such defensive mechanism proposed by the theory is "downwards comparison", whereby individuals recover self-esteem by focusing on comparison with an even lower-status group and emphasizing their own group's positive distinction relative to this new basis of comparison. By derogating an even lower-status group, such as seeking or attending to unfavorable evaluations of it, members of a low-status group effectively extend the perceived hierarchy downward, thereby improving the relative position of their own in-group.[5] Applied to the inter-ethnic context studied here, one might expect the

[5]Relatedly, they might attempt to "sub-type" their group, that is, to emphasize the positive distinction of the sub-group that they belong to relative to the rest of the group (van Bezouw et al. (2021))

established immigrants to respond to the critical reception of their in-group expressed by the native majority by derogating another immigrant group perceived even more critically.

Whereas the SIT explains discrimination downwards as a mechanism of coping with an identity threat resulting from the discrimination received from above, I explore a further (preregistered) mechanisms that could facilitate the effect of received evaluations on participants' attitudes toward refugees.[6] Specifically, I study the role of perceived social norms surrounding the expression of prejudice toward low-status groups in the host society. Previous work on the emergence of social norms shows that individuals, at least in part, infer the group's descriptive norms (what others are doing) from other individuals' behavior to which they are incidentally exposed. In particular, in situations where the behavior of interest does not produce an easily observable outcome (such as discriminatory behavior), people in part rely on direct experiences that they make to learn the descriptive norm regarding this behavior (Kashima et al. (2013), Kwan et al. (2015)). It is thus possible that the groups that were socialized in the presence of a steep ethnic hierarchy and were themselves exposed to prejudiced treatment learn from this experience that discriminating downwards (i.e. against groups ranked in the status hierarchy lower than one's own group) is widespread and perhaps even legitimate behavior in the host society, and are more likely to apply it toward the lower-ranked groups once they encounter them. Accordingly, I hypothesize that exposing low-status immigrants to a negative prejudice directed at their in-group, leads them to perceive expressing negative prejudice toward other (not mentioned) low-status groups (but not high-status ones) as more frequent among the native majority. Should this be the case, it would show that discrimination from above not only can result in an identity threat, which in

[6] One of the applications of the SIT that partly resembles the mechanism studied in this work, although in a different context, is presented by the so-called Queen Bee phenomenon. The term, designates women occupying positions of authority in male-dominated organizations who express a gender bias against women in evaluating their lower-status, young, female subordinates Ellemers et al. (2004). Subsequent work in this literature (for review, see, e.g., Derks et al. (2016)) argues that, rather than being a behavioral trait specific to women, Queen Bee behavior is in itself a response to the gender bias and identity threat in the male dominated environments. Interestingly, the work by Faniko et al. (2017) shows that the negative bias does not apply to other advanced-career women. Instead, the tendency is present only when evaluating a lower-status subgroup of junior female colleagues. Drawing a parallel with the question considered here, one might wonder if there exists a Queen-Bee-Immigrant phenomenon. That is, do the established immigrants respond to encountering a native majority skeptical toward immigrants by expressing a negative bias toward new-coming, lower-status immigrants?

turn motivates discrimination downwards, but it also reduces the perceived social cost of engaging in such behavior through affecting the perceived social norms.

To test this hypothesis, I elicit participants' beliefs regarding the share of the pre-study participants (from the native majority) who negatively evaluated the impact of refugees from the Middle East on the socio-economic and cultural life in Germany. Additionally, to examine whether exposure to prejudice differentialy affects the perceived norm surrounding expression of prejudice toward low-status groups, relative to the high-status ones, I additionally measure participants' expectations of the majority participants' evaluation of several other low-status immigrant out-groups, as well as one high-status out-group. The results provide support for the prediction. Participants who received a negative evaluation of their own in-group expect the native majority to hold significantly more negative views of refugees from the Middle East and of all other low-status out-groups, but not of the high-status one. Whereas the treatment effect on injunctive norms (what others believe one ought to do) is not explicitly tested here, the literature on social norms provides ample evidence for the role that descriptive norms alone play in shaping intentions and behaviors (Bardsley and Sausgruber (2005), Bicchieri and Xiao (2009), Krupka and Weber (2009)) in a wide range of behavioral domains, including expression of prejudice (Álvarez-Benjumea and Winter (2020)).

Finally, beyond the effect on privately held attitudes toward refugees, I explore how exposure to prejudice affects participants' willingness to publicly express them. Previous works studying how privately held opinions translate into publicly expressed attitudes and behaviors found that stigmatization and the social desirability of certain beliefs play an important role in determining the degree to which the discrepancy between the two emerges. In particular, individuals tend to bias their statements when publicly expressed toward positions deemed socially more appropriate (Bursztyn et al. (2018), Enikolopov et al. (2020), Perez-Truglia and Cruces (2017)). Conditional on individuals' private preferences, the importance of their readiness to express them publicly is well demonstrated by the work of Bursztyn et al. (2020). Studying the impact of the rise in popularity of Donald Trump, Bursztyn et al. (2020) show how the public revelation of controversial preferences (such as xenophobic views) can impact the beliefs and behaviors

of the spectators, leading them to be themselves more likely to express and less likely to condemn such attitudes. Therefore, understanding how preference falsification shapes expressed immigration attitudes among established immigrants is not only important as the observable positions might not match the privately held ones, but also because their public expression can be consequential in its own right.

I focus on one aspect of preference falsification and investigate whether established immigrants change their statements when their answers might be observed by a participant from the majority population, and whether this tendency changes with the exposure to prejudice toward their in-group. Participants are asked to provide an answer to the question asking them to rate whether refugees "make Germany a better or a worse place to live" once privately, and once, later in the experiment, after being informed that a future participant, selected from a sample of the majority population, might observe their answer along with the information regarding the participant's region of origin. Comparing the answers provided in both settings reveals that participants indeed do answer differently when their answer is potentially observed, and the direction of misrepresentation depends largely on the initial, privately expressed preference. In particular, participants who provided a more critical assessment of the impact of refugees in Germany when answering privately change their answer toward expressing more supportive views in the observable setting. More interestingly, the opposite holds for the participants who privately assessed the impact of refugees highly positively, that is, they misrepresent their positions in the observable setting so as to appear more critical. Furthermore, among participants who were more critical toward the refugees in the private setting, those assigned to the Negative treatment misrepresent their attitudes in the observable setting (in the positive direction) systematically less than those in the Positive treatment, thus demonstrating the effect of exposure to prejudice on the willingness to express a controversial position publicly.

Given the sustained prominence of immigration as a major issue in political agendas and public narratives across Western countries, the vast body of scholarly work studying immigration attitudes and their determinants comes at little surprise. Most frequently examining the preferences of the native majority, these works identified a set of factors shown to impact the immigration attitudes, including perceived economic threat

(Haaland and Roth (2020), Pardos-Prado and Xena (2019), Facchini and Mayda (2009), Mayda (2006), Scheve and Slaughter (2001)) or cultural threats (Tabellini (2020), Hainmueller et al. (2015), Hainmueller and Hangartner (2013), Card et al. (2012)). This paper contributes to this literature and extends it in at least two directions.

First, I focus on studying immigration attitudes among immigrant population. A small number of existing studies engaging with this question has relied on survey data to either assess average attitudes, or to provide correlational evidence of drivers behind immigrant's views of immigration - including labour market competition, ethnic boundary making and acculturation patterns. To the best of my knowledge this is the first study to consider the effect of immigrants' exposure to prejudice and the resulting status deprivation and causally identify its effect on immigrants' attitudes toward immigration.

Second, this paper goes beyond factors commonly studied in this literature relating to perceived economic and cultural threats of incoming immigration, and examines the importance of the status position occupied by immigrants in shaping their immigration attitudes. In doing so, this work contributes to the literature that examines the importance of individuals' identity and belonging to social groups for their political preferences. The works of Fouka et al. (2021) and Fouka et al. (2020), for instance, engages with a frequently observed setting closely resembling the one studied here – a society populated by a high-status majority and a lower-status minority experiences an entry of a new out-group perceived even more critically. They show how the entry of the new out-group can open a possibility for groups to be re-evaluated and lead the native majority to perceive the old out-group more positively. Examining the attitudes of the established minority toward the new-coming one I take a complementary perspective and show how the negative treatment that the former receives from the majority can trigger defensive behavior leading them to exhibit negative attitudes towards the latter.

Finally, in exploring the mechanism behind the found results, this work contributes to the literature on social norms surrounding expression of stigmatized political preferences. Previous research in this field demonstrates that social norms act as powerful regulator of political behavior, as they guide individuals' perceptions of the degree to which expression of controversial preferences entails social costs

(Álvarez Benjumea (2022), Valentim (2022), Valentim (2021), Bursztyn et al. (2020), Dinas et al. (2020), Barr et al. (2018), Crandall et al. (2002)). For instance, the electoral success of radical political platforms has been shown to effectively signal to individuals that previously stigmatized behaviors, such as e.g. expression of xenophobia or nationalistic views, became (more) socially acceptable (Valentim (2021)), and in turn, increase their propensity to themselves engage in such behaviors (Bursztyn et al. (2020)). I build on this finding and show that even those individuals who are direct targets of respective stigmatized behavior, in present case expression of prejudice, learn from this experience that this behavior is more prevalent and possibly more tolerated in their society.

Immigrants of first and second generation make up an increasing share of the population in most Western societies, and this group is likely to become electorally significant in the following decades. Beyond voting, individuals' views on immigration can shape a range of political behaviors, reaching from protest participation to engagement in violence, as well as shape other political attitudes (Alesina et al. (2023)). Moreover, previous work shows that through interacting with the natives, immigrants' preferences regarding some major political questions can, over time, influence those of the natives and possibly alter future political equilibria (Giuliano et al. (2020)). All of these factors underline the policy relevance of understanding the factors driving immigration attitudes of the immigrant population. This work enters exactly this fray and shows that the common wisdom of immigrants being more likely to support more left-wing platforms (Moriconi et al. (2022)) and holding more positive views on immigration (Berry, 2006; Dandy and Pe-Pua, 2010; Hindriks et al., 2017) cannot be taken for granted, as the latter are malleable and react to contextual factors in systematic ways.

2 Experimental Design

The study is split into two phases, which will henceforth be referred to as the pre-study and the main experiment, both implemented as an online survey. In the following, I provide the description of both phases.

2.1 Pre-study

The pre-study was conducted with a small sample ($N = 125$) of participants residing in Germany and with no immigration background. The only purpose of the pre-study was to collect the responses from the majority population regarding their evaluations of different immigration groups that would later be used in the main experiment.

At the beginning of the survey, participants provided answers to a set of basic demographic questions, including their gender and age, alongside own and parental country (countries) of birth, which were used to ensure that only participants from the majority population with no immigration background, participated in the pre-study.

Thereafter, for each of the several regions/countries, participants were asked to evaluate whether people immigrating from the given region/country contribute rather positively or rather negatively to "socio-economic and cultural life in Germany" (participants selected one of the two options as an answer). To avoid confusion in terms of which countries are encompassed by a given region, with each question participants were shown a simple political map of the relevant part of the world, with the region of interest visibly highlighted, and the text of the question explicitly listing all corresponding countries. Participants in the pre-study were paid a fixed participation fee upon completion of the survey.

2.2 Main experiment

The main part of the experiment was conducted with a sample of 1.159 participants with an immigration background residing in Germany.

Demographics As in the pre-study, at the beginning of the session, participants answered the questions regarding their basic demographic characteristics, including participants' own and parental country of birth. This information was used to match participants to one of the eleven regions of origin[7].

[7]The eligible regions of origin in this study included: Countries in central-eastern European Union (Czech Republic, Slovakia, Poland, Hungary); Romania and Bulgaria; Baltic states (Estonia, Latvia, Lithuania); Countries of ex-Yugoslavia (Bosnia and Herzegovina, Croatia, Kosovo, Montenegro, North Macedonia, Serbia, Slovenia); North Africa (Morocco, Algeria, Libya, Tunisia, and Egypt); Southern European Union countries (Greece, Italy, Portugal, Spain, Cyprus, and Malta); Turkey; Southern countries of the ex-Soviet Union (Tajikistan, Turkmenistan, Georgia, Kazakhstan, Kyrgyzstan, Armenia, and Azerbaijan); Western countries of the ex-Soviet Union (Ukraine, Moldova, Belarus); Russia; and Albania. The division was made with the aim of including the regions of origin most frequently encountered among the population with an immigration background in Germany. At the same time,

Treatment provision In this part of the experiment, participants were informed about the conducted pre-study, in which a group of 125 participants from Germany with no immigration background were asked to evaluate the impact of various immigrant groups on socio-economic and cultural life in Germany, and that some of the collected answers would be shown to them. Participants were then (conditional on the region that they were matched to) randomly split into two treatments. In the **Positive treatment**, participants were shown an answer that evaluated the impact of their own immigrant in-group positively, whereas in the **Negative treatment**, participants were shown an answer that negatively evaluated the impact of their in-group. Here, the positive and negative evaluations refer to the group being evaluated as "contributing rather positively", and respectively as "contributing rather negatively", to socio-economic and cultural life in Germany.

In order to avoid the possibility that participants interpret this information as a signal of a more or less positive attitude toward immigration in general, thus not necessarily reflecting the attitude toward their in-group in particular, the evaluations of two other immigrant out-groups were provided equally in both treatments. In particular, in both treatments, one out-group (immigrants from Western EU countries) was always evaluated positively and the other (immigrants from Lebanon) negatively. Thus, the only difference between the treatments is the evaluation of the own in-group and the inclusion of the two out-groups, consistently evaluated positively and negatively, tying the treatment variation to the position of the own in-group in a (simplified) fixed hierarchy. Figure **??**, provides an example of evaluations presented to participants for both Positive and Negative treatment.

Elicitation of attitudes toward refugees After being presented with the evaluations, participants were told that, in this part of the study, they would be asked to share their opinion regarding immigration to Germany, and in particular regarding persons "currently requesting the right to asylum in Germany". Two measures of participants' support of refugees were then elicited.

The main behavioral measure of participants' support for refugees was captured by the willingness to donate to the United Nations High Commissioner for Refugees (UNHCR).

the division attempted to achieve a trade-off between the number of regions and a sufficiently narrow definition of a region so as to allow for successful clustering.

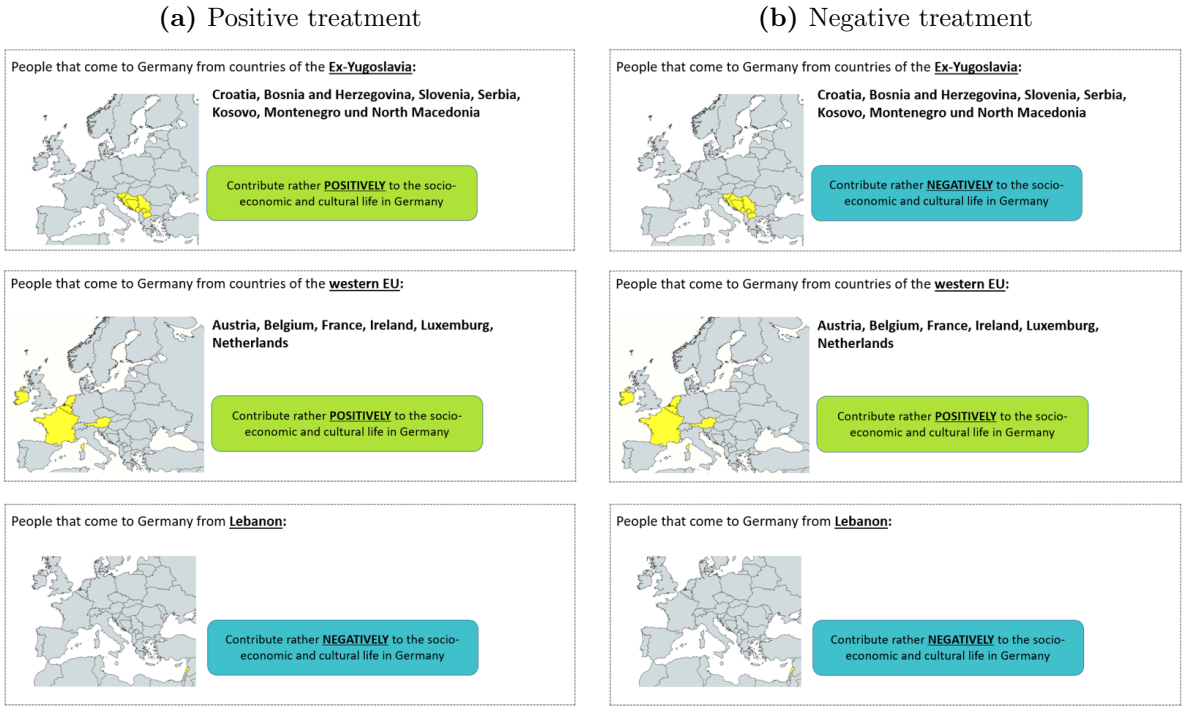

Figure 1. Treatment provision - example

The figure depicts an example of a screen that a participant, whose immigration origin was matched to the region of Ex-Yugoslavia, would see in the treatment provision phase if they were allocated to the Positive treatment (panel a), and that if they were allocated to the Negative treatment (panel b). Participants are informed that they would see a subset of answers collected in the pre-study. Treatment variation is based on randomly matching participants to an answer from the pre-study evaluating the participant's own (parental) region of origin either positively or negatively while keeping the evaluations of the other two out-groups constant.

Participants were informed that, as a part of the study, a lottery would be administered whereby one randomly selected participant would be awarded 100 euros and that all participants would have the same chance of winning the prize. They were then asked whether they would like to donate some part of the 100 euros prize, in the case that they won the lottery, to the UNHCR, which was described to participants as a global organization supporting refugees, and if so, how much. Participants were informed that, if they decided to dedicate some amount to support of refugees, this amount would automatically be deducted from their 100 euro prize in the case they won, and a donation of the same value would be made to the UNHCR.

Moreover, following the approach of Dinas et al. (2021), an attitudinal measure of support was constructed by collecting participants' answers to a set of seven questions. Participants provided their views on (among others) whether Germany should increase

or decrease the number of people it grants asylum to, refugees' influence on the labor market, the welfare state, probability of a terrorist attack, criminality, etc. The list of all questions is provided in the Appendix A.1.

Extended dictator game In this part of the study, participants played an extended version of the dictator game. This allowed for a simultaneous elicitation of two measures of interest. First, the game was designed so as to capture a measure of participants' indirect upstream reciprocity, that is, the tendency of individuals to exhibit prosocial (antisocial) behavior toward others because somebody else has exhibited prosocial (antisocial) behavior toward them. Second, the willingness to share income with an anonymous other participant provided a measure of participants' generosity in an immigration-unrelated context.

Each Participant is assigned one of the three roles: player A, player B, or player C. Thereby, player A is given a budget of 30 euros, out of which they can send a certain sum to another player B, who in turn can send some of the received amount to player C. The amount sent by player A is multiplied by a factor f, and the resulting amount is paid to player B. Player A and player B know that the multiplication factor can take either a high value ($f = 4$) or a low value ($f = 2$), but the realization of this value is not known to any of the players. Thus, player B observes only the resulting sum they receive but is not aware whether it resulted from player A sending a higher sum multiplied by a low factor value, or from player A sending a lower sum multiplied by a high factor value. Here, player A could select between sending 0, 8, 16, and all 30 euros. All participants assigned to role B receive a total of 32 euros (corresponding to player A sending either 8 or 16 euros, and the factor being equal to either 4 or 2, respectively).

Player B is then asked to decide for both scenarios how much of the received sum they would like to send to person C. To ensure that welfare concerns do not play a role in the decision of player B, the amount sent to player C is paid to them without multiplication. Participants are informed that at the end of the study, one triplet will be selected and paid out the amounts according to the decisions they make. Most participants were assigned the role of player B ($n = 1150$), with others distributed among the other two roles.

I take the difference in the amount sent to player C in the scenario where player A

is more generous versus the one when they are less generous as a measure of indirect upstream reciprocity of player B.

Mood elicitation In order to be able to control for the treatments' potential effect on participants' mood, a measure of mood is elicited via the Self-Assessment Manikin questionnaire (Bradley and Lang (1994)). Three questions, intended to capture three major affective dimensions - pleasure, arousal, and dominance - asked participants to select one of the five offered manikins that they feel best describes their mood.

Descriptive norms In order to study treatment effects on participants' perceived descriptive norms regarding the expression of prejudice, in this part of the experiment, participants are asked to guess what percentage of the 125 participants without immigration background that took part in the pre-study evaluated negatively each of several immigrant groups (categorized by their region/country of origin). Each participant is asked to guess the share of participants from the pre-study who negatively evaluated the impact of people immigrating to Germany from: the participant's own (parental) region of origin, Western EU countries (Austria, Belgium, France, Ireland, Luxemburg, Netherlands), Lebanon, Turkey, countries of Southern Africa (South Africa, Namibia, Eswatini and Lesotho) and that of refugees immigrating from the Middle East (Syria, Iraq, Afghanistan, and Pakistan). Participants were informed that the answer closest to the true collected values would be rewarded with an additional 25 euros.

Preference falsification When individuals' are asked to state their political views while observed by the others, preference falsification might mask truly held preferences and skew them to the perceived socially appropriate positions. This part of the experiment aims to capture a potential difference in attitudes expressed by established immigrants when they expect these attitudes to be observed by a majority population, as compared to when this is not the case.

In this part, participants are reminded that all previously provided answers will be delivered only to the researchers in anonymized form. The participants are then informed that only in this part of the experiment are they being asked to provide an answer that can be used in a potential future study to inform future participants about their views on immigration. Furthermore, the instruction clarifies that, if the future study is conducted, it will be run in Germany with a sample of German citizens and that the recipient of

their answer would know their country (countries) of origin. Thereafter participants fill out the answer to the question "Is Germany made a worse or a better place to live by refugees who are granted asylum in Germany?", question that had already been asked as one of the attitudinal questions in the "Elicitation of attitudes" phase.

Additional demographics and debriefing At the end of the experiment, participants are shown the true percentages of participants in the pre-study who negatively evaluated each of the mentioned immigrant groups. The session ended after collecting some additional basic demographic information.

2.3 Data and sample description

The study was conducted in the period December 2021 to January 2022. The sample for the pre-study involved 125 adult individuals with residence in Germany and with no immigration background. A participant was considered to have an immigration background if they, or at least one of their parents, was born outside of Germany. For the purposes of the main experiment, a separate sample was recruited involving 1,176 adult individuals with residence in Germany and with an immigration background. Out of this number, 17 participants provided inconsistent answers to basic demographic questions (e.g., unreasonable age), and their answers were removed, resulting in a sample of 1,159 participants.

Participants with an immigration background were matched to what I refer to for simplicity's sake as "region of origin", indicating one of the eleven regions encompassing their, or parental, country of birth. The regions selected to be targeted in this study encompassed all countries within Europe (except for the Western European countries), all ex-Soviet countries, Turkey, and five northern African countries (Egypt, Tunis, Morocco, Algeria, and Libya). Table A1 in the Appendix A.2 provides an overview of all regions (and all encompassed countries), along with the share of participants matched to each region. The selection of the eligible regions attempted to match the studied sample with the groups most represented among the population with an immigration background in Germany[8], and to focus on those immigrant groups that are more likely to occupy a lower status position than the native majority in German society (thus the exclusion of

[8]See Statistical Office of Germany (Genesis-Online Database, code: 12211-0202)

the Western European countries), while still being (on average) perceived more favorably compared to the refugees from the Middle East. Table A1 in the Appendix A.2 presents the descriptive statistics of the sample across both treatments. The online survey was programmed in Qualtrics and the distribution of the link to the experiment was delegated to a panel company CINT[9].

In the next section, I provide an overview of the empirical results and test the following (pre-registered) hypotheses.

Hypothesis 1 Being assigned to the Negative treatment leads to a decrease in the amount donated to UNHCR and a more negative evaluation of refugees as measured by the attitudinal questions.

Hypothesis 2 Being assigned to the Negative treatment leads participants to expect a higher percentage of negative evaluations of refugees' impact on socio-economic and cultural life in Germany among majority participants (who participated in the pre-study). Furthermore, assignment to the Negative status treatment leads participants to expect a higher percentage of negative evaluation of the own in-group, as well as of the other low-status groups (but not the high-status ones) among majority participants.

Hypothesis 3 Participants with higher indirect reciprocity react more strongly to treatment variation, that is, express more negative (positive) evaluations of refugees in the Negative (Positive) treatment.

Hypothesis 4 The distribution of answers provided to the question "*Do refugees who obtain the right to asylum in Germany make Germany a worse or a better place to live*" in the "private" scenario differs from the distribution of answers provided to the same question in the "observable" scenario. Furthermore, being assigned to the Negative treatment leads participants to express a less favorable opinion of refugees in the "observable" scenario.

3 Results

3.1 Pledged donation to the UNHCR

In this subsection, I present the measured effect of the treatment, that is, the effect of receiving negative status information, compared to receiving positive status information,

[9]https://www.cint.com/

on the behavioral measure of participants' support for refugees. The measure of support is captured by the amount that participants committed to donate to the United Nations High Commissioner for Refugees (UNHCR), from a 100-euro prize that is raffled among all participants at the end of the study. On average, participants committed to donate 36.86 euros, with individual decisions spanning across the full range of possible donations.

The results presented in Table 1 depict the effect of being allocated to the Negative treatment (with the Positive treatment serving as a baseline) on participants' pledged donations to the UNHCR. Considering that the possible value of the donation was limited at 0 from below, and at 100 from above, and that the number of participants who selected both limiting values was significant, the table presents the results of a Tobit regression of the donated amount on the treatment variable and individual controls[10]. All presented regressions include fixed effects of the federal state within Germany and the region of participants' (parental) origin.

The results in Table 1 provide support for Hypothesis 1. The results shown in column (1) demonstrate that participants in the Negative treatment committed to donate systematically less to the UNHCR (relative to those allocated to the Positive treatment). Participants pledged to donate on average around 4.7 euros less if they were in the Negative treatment ($p < 0.01$), corresponding to around 13% of the average pledged donation. Furthermore, as shown in column (3), participants allocated to the Negative treatment were also significantly less likely to pledge any positive donation. In particular, reallocating a participant from the Positive to Negative treatment decreased, on average, the probability of the participant pledging a positive donation by 5.4 percentage points ($p < 0.01$). The results in columns (2) and (4) show that these findings are robust to the inclusion of controls for the respondents' socio-demographic background.[11]

One concern that might arise here, considering the method of delivering the treatment variation, is the possibility that receiving a stark negative (positive) evaluation message regarding participants' in-groups could simply set them in a bad (good) mood and, through this, impact their prosociality more generally. In this case,

[10]The OLS analysis produces qualitatively same results and is depicted in Table A2 in Appendix A.3.

[11]Table A2 in the Appendix shows the effect of the demographic characteristics on the pledged donation. Other than age, which has a negative effect on the pledged donations, none of the other demographic measures had a significant effect.

Table 1. Treatment effects: Pledged donation to the UNHCR

	(1)	(2)	(3)	(4)
	Pledged donation		Pr(Donation>0)	
Negative treatment	-7.049***	-6.922***	-0.189***	-0.190***
	(1.593)	(1.532)	(0.063)	(0.060)
Constant	47.418***	54.824***	1.031***	1.400***
	(3.082)	(4.608)	(0.152)	(0.212)
Marginal effects: $E(\Delta y/\Delta x)$				
Negative treatment	-4.724***	-4.637***	-0.054**	-0.054**
	(1.060)	(1.022)	(0.018)	(0.017)
Individual controls	No	Yes	No	Yes
Observations	1,159	1,159	1,159	1,159

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Column (1) and column (2) show the results of the Tobit regression of the amount that a participant pledged to donate to the UNHCR on the treatment variable and the set of individual controls. Negative treatment indicates receiving a negative evaluation of the own in-group (with Positive treatment serving as a baseline). Reported marginal effects represent the average marginal effect of being allocated to Negative treatment on the pledged donation value. Columns (3) and (4) show the results of Probit regression of the dummy variable that takes value one if a participant pledged to donate a value larger than zero on the treatment variable and the set of individual controls. All regressions include fixed effects of the federal state of residence in Germany and the region of participants' (parental) origin. Individual controls (included in columns (2) and (4)) include age, gender, equivalent household income tertile, and indication of tertiary education. Reported marginal effects represent the average marginal effect of being allocated to Negative treatment on the probability of pledging to donate any positive amount and can be directly interpreted in terms of percentage points difference. Standard errors in parentheses are clustered on the level of the region of participants' (parental) origin.

the observed difference in pledged donations would not necessarily reflect the specific change in attitude toward other immigrant out-groups. In order to control for this possibility, I compare participants' giving in a dictator game, in which they decide, for two possible scenarios, the amount they want to send to an anonymous other participant[12]. The results from this game are discussed in more detail in section 3.4. Here, it suffices to say that the dictator giving in both treatments was remarkably similar and, if anything, slightly higher in the Negative treatment (both-sided t-test on the equality of means: $p > 0.5$ in both scenarios), providing evidence that the findings depicted in Table 1 cannot be explained by changes in participants' broader generosity.

[12]Participants had no knowledge about the criteria that were used to select individuals eligible to take part in the Experiment, and thus had no information on the region of origin, or even on the immigration status of the participant that they would be matched with.

A further test is provided by comparing the collected measures of participants' mood across treatments. However, the distribution of all three measured affective dimensions - pleasure, arousal, and dominance, elicited via Self-Assessment Manikin questionnaire (Bradley and Lang (1994)), did not differ significantly between the two treatments (Kolmogorov–Smirnov test for equality of distribution in both treatments, for each of the three affective dimensions - pleasure: $p > 0.6$; arousal: $p > 0.9$; dominance: $p > 0.9$).

3.2 Attitudinal measures

In addition to the behavioral measure of support for refugees, a set of attitudinal measures was elicited by means of collecting answers to seven questions regarding refugees requesting asylum in Germany. The questions, among others, regraded participants' views of the influence of refugees on employment, risk of terrorism, criminality. The exact formulation of all seven questions is provided in Appendix A.1.

Table 2. Treatment effects: Attitudinal questions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	q1	q2	q3	q4	q5	q6	q7	$\bar{q}$
Negative treatment	-0.266**	-0.082	-0.051	-0.031	0.043	-0.144	-0.049	-0.107
	(0.109)	(0.105)	(0.122)	(0.111)	(0.126)	(0.089)	(0.040)	(0.105)
Constant							1.082***	2.120***
							(0.174)	(0.178)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,159	1,159	1,159	1,159	1,159	1,159	1,159	1,149

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Columns (1) through (6) show the results of the ordered logistic regression of participants' answers responding to the respective 6 attitudinal questions on the treatment variable and the set of individual controls. All answers are re-coded such that a higher value indicates higher support for refugees. Column (7) shows the result of a Probit regression of a dummy variable that takes value 1 if a participant selected "To flee war" or "Avoid political persecution" as the primary reason why refugees leave their countries, and 0 otherwise. All regressions include fixed effects of the federal state of residence in Germany and the region of participants' (parental) origin. Individual controls include age, gender, equivalent household income tertile, and indication of tertiary education. Standard errors in parentheses are clustered on the level of region of participants' (parental) origin.

Compared to the effect on pledged donations, treatment variation had a smaller effect on the attitudes reported in the seven questions. The first six columns of Table 2 show the results of an ordered logistic regression of the chosen answer for each of the (first six) questions on the treatment variable and the set of socio-demographic controls. All answers

are re-coded such that a higher value indicates higher support for refugees. Column (1) shows that in the case of the first question, which asked participants' opinion on whether Germany should increase or decrease the number of people it grants asylum to, participants were significantly more likely to provide a lower answer (decrease the number of granted asylums) if they were in the Negative treatment. However, although treatment effects work in the predicted direction in most of the other questions (that is, participants in the Negative treatment provided less supportive answers), these effects are not significant. Thus, while affecting the overall attitude toward refugees, as reflected by the pledged donations and the desired level of entry that should be allowed, treatment variation did not change how participants perceive particular topical concerns frequently discussed in the context of the immigration of refugees.

Question q7 asked participants to provide their opinion on the primary reason why refugees abandon their countries among the following options: "to flee war", "avoid political persecution", "improve their economic conditions" and "obtain access to social security payments in the destination country". I construct a dummy variable that takes value one if a participant selected one of the first two choices and show in column (7) the results of the Probit regression of this variable. Again here, being assigned to the Negative treatment decreased the probability of selecting one of the two reasons that would indicate security (rather than economic) concerns as a primary reason for flight, but the effect is insignificant.

Finally, I construct an aggregate measure of participants' answers to attitudinal questions by averaging for each participant seven dummy variables. The dummy variables correspond to the seven questions, and each takes value one if the participant selected an answer to the respective question that indicates higher support for refugees than that implied by the neutral point (selected 3 (resp. 5) on a scale of 1 to 5 (resp. 0 to 10)). Column (8) shows the results of regressing this aggregate measure, denoted by $\bar{q}$, on the treatment variable and the set of individual controls.

Another point of caution is worth discussing here. Namely, the statements of evaluation of immigrant groups that were provided to the participants were purposefully formulated very broadly so as to deliver a prejudiced and generalizing evaluation of the groups without pointing out any particular characteristic that might

be ascribed to them[13]. Nevertheless, one might still be concerned that the statement in the Negative treatment could affect how participants perceive the socio-economic standing of their in-group, and thereby possibly reinforce (or weaken) the perceived threat of intensified labor market competition, or the competition for welfare benefits, that might result from the immigration of refugees (and the other way around in the Positive treatment). However, the answers to question q2, which asked the participants whether the refugees "take away our jobs and social benefits", did not significantly differ between the treatments, thus providing evidence against this possibility. Furthermore, interacting the treatment variable with the measure of income, or with the indication of higher education and adding them to the regression of the pledged donations (Table 1, column (2)), does not qualitatively change the found results, and the coefficients of the interactions remain insignificant (the results of these regressions are presented in Table A3 in Appendix A.4). In other words, taking education and household income as proxy measures of participants' vulnerability to the competition from the newly incoming immigration, it does not seem to be the case that the exposure affects the average pledged donation (in any of the treatments), nor that those with a higher exposure differentially react to the treatment variation.

Finally, should the evaluations manipulate participants' view of their own economic standing, and through this affect their perception of whether they can afford (or should be expected) to help others in general, this difference between the treatments should also be reflected in the dictator giving, which was not the case.

3.3 Descriptive norms regarding the expression of prejudice

Results in the previous section showed that exposing participants to a negative evaluation of immigrants from their own (parental) region of origin led them to significantly decrease their support for refugees. As discussed above, Social Identity Theory offers an explanation of such behavior among low-status groups as a means of coping with the identity threat stemming from the low-status position they occupy. Individuals from such groups might be tempted to recast their group identity in a more positive light by emphasizing its positive distinctiveness relative to some lower-ranked

[13] As a reminder, the statements read: People immigrating to Germany from (a given region) contribute (rather positively) / (rather negatively) to socio-economic and cultural life in Germany.

group. This effectively extends downward the relevant hierarchy, and thus improves the relative position of their in-group.

On the other hand, the cost of engaging in this strategy will arguably depend on the societal context. Emphasizing the positive distinctiveness of one's own (immigrant) ethnic/national in-group relative to another frequently relies on employing generalizations and ethnic or racial stereotypes, and the legitimacy of the use of these concepts is significantly shaped by the social norms prevailing in the relevant environment (Álvarez Benjumea (2022), Bursztyn et al. (2020), Barr et al. (2018), Crandall et al. (2002)). Therefore, the readiness of the immigrant groups to respond to an identity threat by engaging in downward competition can be expected to depend on the perceived norms surrounding the expression of prejudice toward low-status groups in the host society.

Interestingly, individuals exposed to prejudice from the majority population might use this experience to update their perception of the norms regulating expressions of prejudice in the host society. In other words, low-status immigrants could learn from discrimination directed toward their own in-group that discriminating downwards (i.e. against groups ranked lower than one's own group) is a widespread and possibly also acceptable behavior in the host society. The prejudice encountered by low-status immigrants could therefore work not only as an identity threat motivating downward competition, but also so as to reduce its perceived social cost.

In the context of the conducted experiment, this would imply that participants rely on the received evaluation of their in-group to update their expectation of the acceptance expressed by the native majority, not only toward their own, but toward all lower-status groups, including refugees from the Middle East (as proposed by Hypothesis 2). In order to test this prediction, I collect an incentivized measure of the descriptive norm regarding prejudice expression toward low-status groups prevailing in the host society. To facilitate norm elicitation, participants were asked to guess the share of respondents in the pre-study (without migration background) who evaluated *negatively* the impact of each of several immigrant groups on socio-economic and cultural life in Germany. Particularly, each participant was asked to guess the share of the native majority participants who negatively evaluated the impact of refugees

immigrating from the Middle East. Additionally, the same estimation question was asked regarding the evaluation of people immigrating to Germany from Turkey, Lebanon, Southern Africa and participants' own (parental) region of origin. Finally, in order to check whether exposure to prejudice (stemming from a higher-status majority) differentially impacts participants' expectations of expression of prejudice toward low-status groups (relative to the high-status ones), participants were asked the same question regarding the evaluation of the immigrants coming from the countries of Western European Union. Recall that in the treatment provision participants in both treatments saw a positive evaluation of the immigrants from Western European Union and a negative evaluation of immigrants from Lebanon, while the other groups were not mentioned in any way until this part of the experiment. The exact phrasing of the question and an example screen seen by participants is provided in the instructions available in online Appendix[14].

Table 3, provides an overview of measured treatment effects on participants' expectations. The results in the first column show that participants who received a negative evaluation on their own in-group, on average, expected the majority population participants to be more critical toward immigrants from their region of origin. This is also intuitive, as it reflects the information that participants received in the treatment provision, but is still informative as it shows that participants extrapolated from the individual evaluation they received to the average opinion of the group. At the same time, it serves to confirm the successful treatment manipulation.

More interestingly, the same applies to participants' expectations of evaluations of all other low-status immigrant groups. Particularly, in accordance with Hypothesis 2, participants in the Negative treatment expected a significantly more negative evaluation of the impact of refugees from the Middle East, as well as of people immigrating from Turkey, Lebanon and from countries in the south of Africa. This is not the case for the expected evaluation of high-status immigrants, that is, those coming to Germany from Western EU countries, indicating that this is not a consequence of expecting the majority population to be more sceptical toward immigrants in general. Instead, as proposed by Hypothesis 2, it appears that receiving a negative evaluation of their own in-group led

[14]Full survey instructions are available at http://biljanameiske.com/wp-content/uploads/2022/10/Queen-Bee-Immigrant_Instructions_ENG.pdf

participants to expect more critical views only of those immigrant groups that were of a lower status than those who are providing the evaluation.

Table 3. Treatment effects: Empirical expectations

Elicited expectation:	What percentage of majority population participants evaluated negatively the impact of people coming to Germany from:					
	(1)	(2)	(3)	(4)	(5)	(6)
	Own (parental) region of origin	Refugees	Turkey	Lebanon	Countries in Southern Africa	Western EU countries
Negative treatment	5.863***	4.921**	3.524**	4.184*	5.112*	-1.463
	(1.118)	(2.187)	(1.161)	(2.021)	(2.444)	(1.330)
Constant	50.237***	53.229***	50.914***	38.812***	37.494***	41.056***
	(3.080)	(4.129)	(4.560)	(4.277)	(3.734)	(2.375)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,159	1,159	897	1,159	1,159	1,159

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Results of the OLS regression of the elicited guess (of participants with an immigration background) of the share of majority population participants who evaluated negatively the impact on socio-economic and cultural life in Germany of people immigrating from countries/regions shown in columns' headers. The first column regards the evaluation of the people immigrating to Germany from the country/region of participants' origin (or that of their parent(s) if the participant was born in Germany). In questions that regarded immigrants from a region (rather than a country) all countries within the region were listed in the question. All regressions include fixed effects of the federal state of residence in Germany and the region of participants' (parental) origin. Standard errors are clustered on the level of region of participants' (parental) origin. Individual controls include age, gender, equivalent household income tertile, and indication of tertiary education.

While the literature on social norms provides evidence of an impact of descriptive norms (what others are doing) on injunctive norms (what others believe one ought to do) (see e.g. Bicchieri et al. (2020b), Bicchieri et al. (2020a)), the results provided here can only support the treatment effect on the former. On the other hand, irrespective of their influence on the injunctive norms, descriptive norms have been shown to, on their own, influence behavior in a wide range of domains (Krupka and Weber (2009), Bardsley and Sausgruber (2005)). Particularly in value-laden domains, where, due to social pressure, the injunctive and descriptive norms might be in conflict, the perception of the descriptive norms was found to be the one predicting the behavior (Bicchieri and Xiao (2009)). Whereas the ultimate test for the behavioral effects of the descriptive norms would amount to administering a norm-manipulation experiment and is thus outside the

scope of this work, these results suggest that the observed treatment effect on descriptive norms regarding the expression of prejudice toward refugees, can be expected to affect participants' pledged donations to the UNHCR.

3.4 Indirect reciprocity

Another reason behind the effect that receiving evaluation on own (parental) region of origin had on support for refugees might be upstream indirect reciprocity. Upstream indirect reciprocity designates a tendency of individuals to exhibit prosocial behavior toward others because somebody else has exhibited prosocial behavior toward them (Alexander (1987), Nowak and Sigmund (2005)). Previous studies have provided evidence for upstream indirect reciprocity, both in the laboratory (Greiner and Levati (2005)) and in field experiments (Mujcic and Leibbrandt (2018)). Exhibiting upstream indirect reciprocity in the context of inter-minorities relations would suggest that receiving a less (more) favorable evaluation from an out-group might translate into a less (more) favorable view of another out-group. Thus we would expect more reciprocal participants to react more negatively (positively) in terms of their support for refugees if they were assigned to the Negative treatment (Positive treatment).

In order to test this prediction, a measure of indirect upstream reciprocity was collected using an extended dictator game, whereby one participant (player A) can send a certain sum to another participant (player B), who in turn can send some share of the received amount to a third participant (player C). The amount sent by participant A is multiplied by a factor, which can take either a high or a low value, but the realization of this value is not known to any of the players. Thus, player B observes only the resulting sum they received but is not aware whether it resulted from player A sending a higher sum that was multiplied by a low factor value, or from player A sending a lower sum that was multiplied by a high factor value. Player B is then asked to decide for both scenarios how much of the received sum they would like to send to person C. To ensure that welfare concerns do not play a role in the decision of player B, the amount sent to player C is paid to them without multiplication. Each participant is matched to one of the three roles, and a randomly selected triplet is paid out the amounts according to the decisions they made.

I take the difference in the amount sent to player C in the scenario where player A was more generous versus when they were less generous as a measure of indirect upstream reciprocity of player B. In order to collect this measure for as many participants as possible, most of the participants were assigned the role of player B ($n = 1150$), and the rest was distributed among the other two roles. All participants assigned to role B received a total of 32 euros (corresponding to player A sending either 8 or 16 euros, and the factor being equal to either 4 or 2, respectively). On average, participants sent 1.21 euros more to player C when player A sent them a higher amount compared to when they sent a lower amount (average amounts sent in two cases were 13.99 and 12.79 euros). This difference is significant (Wilcoxon signed-rank test: $z = 9.544$, $p < 0.001$), providing evidence for behavior consistent with indirect upstream reciprocity. Furthermore, the distribution of the measure of indirect reciprocity does not differ among treatments (Kolmogorov–Smirnov test: $p = 0.785$), supporting the view of reciprocity as a basic preference.

Table 4. The role of upstream indirect reciprocity

	(1)	(2)
	Donation	Pr(Donation>0)
Negative treatment	-5.815***	-0.139*
	(1.557)	(0.073)
Ind. reciprocity	0.433	0.009
	(0.267)	(0.006)
Negative treatment*Ind. reciprocity	-0.751*	-0.033**
	(0.446)	(0.015)
Constant	46.482***	1.002***
	(3.089)	(0.152)
Individual controls	No	No
Observations	1,150	1,150

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Column (1) shows the results of the Tobit regression of the amount that participants pledged to donate to the UNHCR on the treatment variable, a measure of upstream indirect reciprocity (denoted Ind. reciprocity), and their interaction. Column (2) shows the results of the Probit regression of the dummy variable that takes value one if a participant pledged to donate a value larger than zero on the same set of regressors. All regressions include fixed effects of the federal state of residence in Germany and the region of participants' (parental) region of origin. Standard errors are clustered on the level of participants' (parental) region of origin.

Table 4 provides the results of the regression of the amount pledged to donate and that of the dummy variable indicating that the participant pledged a positive donation

on the treatment variable, measure of indirect reciprocity and their interaction. The results indicate that indirect reciprocity indeed had some role in determining the decision to donate. Whereas indirect reciprocity in the Positive treatment increased, albeit insignificantly, the pledged donation ($coef. = 0.433$, $p = 0.105$) and the probability to donate ($coef. = 0.009$, $p = 0.167$), it significantly reduced both values in the Negative treatment. However, although providing some evidence for the role of indirect upstream reciprocity, these effects are relatively small and do not provide a systematic explanation of the found treatment effects (the treatment variable remains significant).

3.5 Preference Falsification

Previous subsections aimed to describe how exposure to expressed prejudice shapes the immigration attitudes of individuals with an immigration background when these attitudes are expressed privately, that is, when they are unobservable to others (other than the experimenter). However, a broad range of political behaviors, such as protesting, signing a petition, or publicly expressing political views, are per construction observable to other members of the polity, and as such are susceptible to social effects. In particular, due to perceived social pressure, individuals with counter-normative views may prefer to falsify them under observation (Kuran (1997)), such that expressed preferences might not always fully match privately held ones. Previous empirical works demonstrate convincingly that individuals care about how they are perceived by others, and that reputational concerns consequently shape observable behavior in a variety of settings, including political behavior (Valentim (2022), Bursztyn et al. (2020), Dinas et al. (2020), Enikolopov et al. (2020), DellaVigna et al. (2016), Gerber et al. (2008)).

Understanding how perceived social pressure in the host society might impact the expressed immigration attitudes of established immigrants is important not only because preference falsification might mask their genuine preferences but also in light of the findings that the expression of controversial preferences, such as xenophobia, might have far-reaching spill-over effects on the beliefs and behaviors of others who observe it, by e.g., leading them to be themselves more likely to express and less likely to condemn such attitudes (Álvarez-Benjumea and Winter (2020), Bursztyn et al. (2020), Bursztyn

et al. (2018)). In the extreme case, this might even lead to the unraveling of norms that protected against the respective behaviors and attitudes. Therefore, understanding factors that facilitate public expression of controversial preferences conditional on their existence is important in its own right.

In the context observed here, I focus on one possible mechanism that might lead to preference falsification and study whether established immigrants change expressed preferences toward refugees when these preferences are potentially observable by the members of the native majority (and, if so, in which direction). Furthermore, I analyze whether being exposed to the expression of prejudice toward their in-group has an effect on participants' willingness to misrepresent their attitudes toward the refugees under observation.

To get some insight into this, one of the questions that was used in collecting the attitudinal measure of support for refugees (q6) was asked again later in the survey, but participants were this time informed that their answer might be shown to a participant in a future study. Participants knew that, if used, their response would be provided to a future participant (in an anonymized form) along with an indication of whether the participant had an immigration background and, if so, from which country, and that the person observing their answer would be a German citizen. The question asked participants to rate whether refugees who obtain the right to asylum in Germany make Germany a worse or a better place to live. Participants answered by selecting a number on an 11-points number line, where 0 was indicated as "worse place to live", and 10 as "better place to live". Note that participants were given the opportunity to provide a neutral answer by selecting 5 on the number line, which is exactly in the middle between the two extremes.

I denote the two scenarios as "private" and "observable"[15], and the answers provided in them by a_p and a_o, respectively (note that a higher answer indicates a more supportive attitude toward refugees). To compare the answers provided in the two scenarios, I construct a variable $\Delta_o = a_o - a_p$, capturing the extra support that participants expressed in the observable scenario relative to that in the private scenario.[16]

[15]The use of the terms "observable" and "private" here is intended only to designate and make easier the distinction between the two scenarios. The ability of the researchers to observe participants' answers renders the private setting clearly distinct from a truly private setting.

[16]I argue that calculating a difference, in this case, is appropriate as the question explicitly asked the

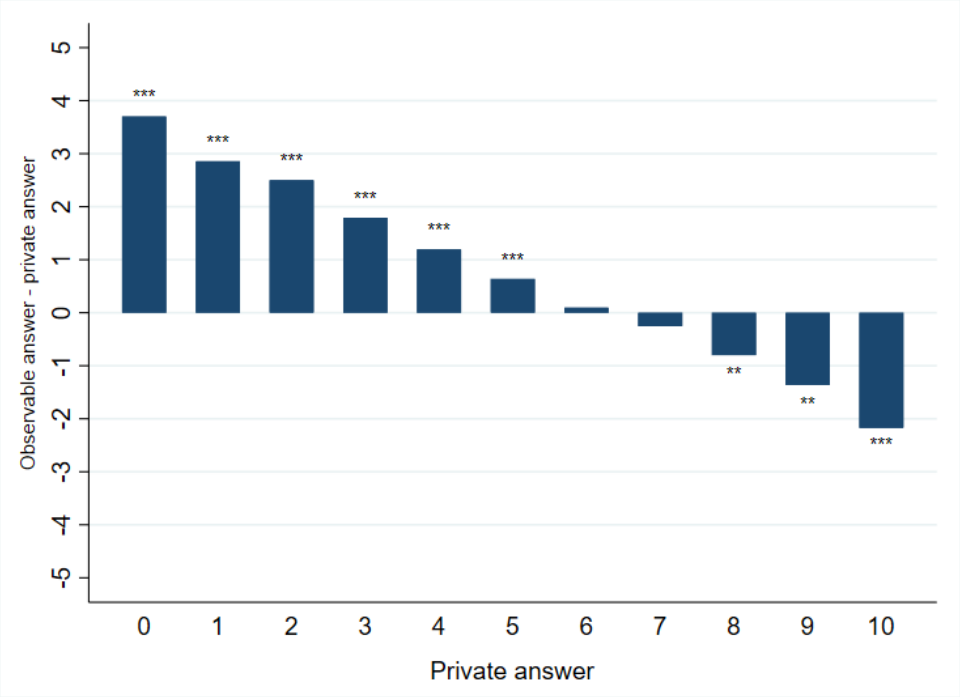

Figure 2. Difference in expressed attitudes - observable vs. private scenario

The figure depicts the average preference falsification, captured as a difference between the answers provided in the "observable" and in the "private" scenario ($\Delta_o = a_o - a_p$). The average difference between the answers is depicted per answer provided in the private scenario. A positive (negative) value indicates that, on average, participants provided an answer implying a more (less) supportive attitude toward refugees when their answer will possibly be observed by a future participant (German citizen) than when answering privately. Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ in sign test for $H_0 : median(\Delta_o) = 0$

Figure 2 shows the average value of Δ_o over a_p. The figure indicates that the average difference in answers strongly depends on the value of the answer initially provided in the private scenario. In particular, participants who expressed less support in the private scenario (provided any answer up to the neutral point (5)), on average, provided systematically higher answers in the observable scenario. More interestingly, participants who in the private scenario indicated highly supportive attitudes ($a_p > 7$) systematically decreased their answers in the observable scenario. This suggests that established immigrants, when given an opportunity to misrepresent their attitudes in front of the majority population, do not only use it so as to present themselves as more

participants to rate refugees' influence on a visibly enumerated line, with only endpoints carrying the (exactly opposite) labels. As the answer options are number values (rather than statements, as would be the case in a standard Likert scale with different levels of agreement), collected answers can be considered as interval data.

tolerant than they are , but also to present themselves as less tolerant than they truly are.[17]

The results depicted in Table 5, illustrate the effect of experimentally induced status on preference falsification. The table shows the results of an OLS regression of the measured preference falsification (Δ_o) on the treatment variable, while controlling for the privately expressed preference (a_p) and a set of individual characteristics. In order to account for the heterogeneous response to treatment across the distribution of the privately expressed preference, I run the regression separately for participants expressing different levels of support in the private scenario. Particularly, columns (1), (2), and (3) include participants who, in the private scenario, chose an answer that indicates (increasingly) more critical views than one that would be indicated by selecting a neutral point at $a_p = 5$. Accordingly, columns (4), (5), and (6) include participants who privately indicated (increasingly) more supportive attitudes.

The results show that, among participants who privately indicated more critical attitudes ($a_p < 5$, column (1)), being allocated to the Negative treatment systematically reduced preference falsification. In other words, whereas critical participants falsify their attitudes so as to appear more tolerant in both treatments, those allocated to the Negative treatment do so significantly less. The treatment effect increases in size and precision among those who expressed even more critical views privately ($a_p < 4$, column (2) and $a_p < 3$, column (3)). On the other hand, assignment to the Negative treatment (while still having a negative sign) did not significantly affect preference falsification among those who privately expressed attitudes that are more supportive than that indicated by the neutral point (i.e., for whom $a_p > 5$), neither when observed together (column (4)), nor when focusing only on those with particularly supportive views (column (5) and column (6)).

These results suggest that expressed prejudice not only negatively affects the privately held attitudes toward refugees of those exposed to it, but also increases the readiness to publicly present biased views, thereby weakening the effect of the social norm against xenophobic expressions.

[17] One concern here is that the presented evidence of mean reversion when comparing answers in private and observable scenarios might have also resulted if the participants randomly selected their answers in both cases. I discuss this possibility in Appendix A.5 and present the evidence against this case.

Table 5. Treatment effects: Preference falsification

	(1) $a_p < 5$	(2) $a_p < 4$	(3) $a_p < 3$	(4) $a_p > 5$	(5) $a_p > 6$	(6) $a_p > 7$
			$\Delta_o = a_o - a_p$			
Negative treatment	-0.435*	-0.661**	-1.154***	-0.063	-0.165	-0.157
	(0.219)	(0.234)	(0.283)	(0.232)	(0.285)	(0.431)
a_p	-0.650***	-0.667***	-0.720**	-0.594***	-0.690***	-0.826***
	(0.104)	(0.137)	(0.276)	(0.082)	(0.124)	(0.164)
Constant	4.088***	4.050***	5.098***	4.102***	4.985***	5.821***
	(0.267)	(0.403)	(1.097)	(0.913)	(0.987)	(1.487)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	394	288	184	510	395	265

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Results of the OLS regression of the measure of preference falsification $\Delta_o = a_o - a_p$, on the privately provided answer a_p and treatment variable. Columns (1) through (3) include only those participants who in the private scenario chose an answer that indicates (increasingly) more critical view than the one that would be indicated by selecting a neutral point at $a_p = 5$, as indicated in the columns' title line. Conversely, columns (4) through (6) include only those participants who in the private scenario chose an answer that indicates (increasingly) more supportive view than the one that would be indicated by selecting a neutral point at $a_p = 5$. All regressions include fixed effects of the federal state of residence in Germany and the region of participants' (parental) region of origin. Standard errors are clustered on the level of participants' (parental) region of origin.

4 Conclusion

While immigration attitudes have received a lot of attention in both economics and political science literature, previous research predominantly examined these positions from the point of view of the majority populations of receiving countries. This paper studies immigration attitudes of established immigrants, that is, those who have already resided in the host countries for some time, toward new flows of immigration, and the drivers behind these positions. Starting from the implications of Social Identity Theory, I hypothesize that relative status deprivation, that is, the negative difference in status between own ethnic/national group and that of the native majority (or other, more favorably perceived minorities), has a negative impact on group's members' attitudes toward an even lower ranked status group (e.g., such as refugees). I argue that low-status groups that were socialized in a steep ethnic hierarchy and were exposed to prejudice, over time come to perceive ethnic competition as usual and perhaps legitimate, and consequently engage in it also when they are faced with even lower-status groups.

In an online experiment, a sample of participants with an immigration background residing in Germany is randomly assigned to receive either a positive or a negative evaluation of the influence of their own (immigrant) in-group on "socio-economic and cultural life in Germany", as expressed by a participant from the majority population (with no immigration background). Participants are additionally provided with the evaluations of two other out-groups (same for all participants), which fixes the status hierarchy and only leaves the position of the participant's in-group variable. Experimental results confirm the hypothesis by showing that participants who received a negative evaluation of their in-group are significantly less willing to pledge a donation to the UNHCR, and provide less supportive answers to a set of questions regarding attitudes toward refugees (albeit the latter difference is only partially significant).

Furthermore, I hypothesize that a part of the detected effect of the exposure to prejudice works through manipulating the social norms surrounding discrimination and its expressions. In particular, people from low-status regions could learn from discrimination directed toward their own in-group that discriminating downwards (i.e. against groups ranked lower than one's own group) is a widespread behavior in the host society, which in turn increases the probability of them engaging in such behaviors themselves. The results show that when asked to guess how participants from the native majority evaluated the impact of other immigrant groups, participants who received a negative evaluation of their own in-group (compared to those who received a positive one) expect the evaluations to be significantly more critical. This applies to the expected evaluations of all (mentioned) low-status immigrant groups, including the refugees from the Middle East, but not to the evaluation of a high-status immigrant group. I provide tentative evidence for the role of perceived descriptive norms regarding the acceptance of refugees in mediating the treatment effect on behavior.

Lastly, I show that receiving a negative evaluation of the in-group increases the readiness of those participants who privately hold the most negative attitudes toward refugees to publicly state their views, thus weakening the effect of the norm against xenophobic expressions.

The findings presented in this work show how factors specific to the receiving, rather than sending country, might impact the political views and behavior of immigrants.

They highlight the importance of policies and public attitudes affecting perceptions of the immigrant groups' status, and particularly those seeking to regulate expressions of prejudice, by showing how status effects spill over into attitudes toward other (and perhaps not yet present) minorities.

A Appendix

A.1 Attitudinal questions on views regarding refugees from the Middle East

Participants were asked to provide answers to the following seven questions: 1) Do you think Germany should increase or decrease the number of people it grants asylum to? (1 = Greatly increase; 5 = Greatly decrease); 2) Refugees are a burden on our country because they take our jobs and social benefits.(1 = Completely agree; 5 = Completely disagree); 3) The money spent on the accommodation of refugees in our country could have been spent better to cover the needs of Germans. (1 = Completely agree; 5 = Completely disagree); 4) Refugees will increase the likelihood of a terrorist attack in our country. (1 =Completely agree; 5 = Completely disagree); 5) Refugees in our country are more to blame for crime than other groups. (1 =Completely agree; 5 = Completely disagree); 6) Is Germany made a worse or a better place to live by refugees who are granted asylum in Germany? (Respondents select their answer on a enumerated scale, where value 0 is labeled as "Worse place to live", and value 10 is labeled as "Better place to live"); 7) Among the following options, which one do you think best explains why refugees from Syria and other countries leave their country? (1 = To flee war; 2 = To improve their economic conditions; 3 = To avoid political persecution; 4 = To gain access to host country's social benefits.)

Other than question number 6, all questions have been adopted from Dinas et al. (2021). For the purposes of the analysis presented in 2 all answers were re-coded such that a higher value indicates higher support for refugees.

A.2 Sample description

Table A1 shows the distribution of the sample across the regions of origin and the basic demographic characteristics of the sample.

A.3 OLS analysis of the pledged donation amount

The following table depicts the results of the OLS regression of the amount that participants pledged to donate to the UNHCR on the treatment variable and the set of individual controls. The results corroborate the findings presented in Table 1.

A.4 The role of participants' socio-economic characteristics

Table A3 restates the results of the regression of participants' pledged donation to the UNHCR as depicted in Table 1, but with the addition of controlling for the interaction of the treatment variable with the measure of income (column (1)), and its interaction with the indicator of participants having completed any post-secondary education.

A.5 Average preference falsification

One concern regarding the presented results on average preference falsification is that the presented evidence of mean reversion when comparing answers in private and observable scenarios (Figure 3) might have also resulted from participants randomly selecting their answers in both cases. However, multiple findings suggest that this is unlikely the case. Firstly, the distributions of answers in both scenarios, a_p and a_o, both significantly differ from the uniform distribution (Kolmogorov-Smirnov tests for $a_p = U(0, 10)$, and for $a_o = U(0, 10)$, both reject the null hypothesis with $p < 0.001$). Furthermore, the answer to the question in the private scenario is significantly correlated both with the answer in the observed scenario, as well as with the answers to all other attitudinal questions (coefficient of correlation between 0.40 and 0.47 and $p < 0.001$ in all pairwise tests) and to the donation (coef.=0.23, $p < 0.001$), suggesting that participants did not answer the question at random. Finally, as evident from Figure 3, the observed degree of preference falsification is significantly lower than the one expected if participants had answered randomly in both scenarios.

Nevertheless, this does not exclude the possibility that some share of participants randomly selected their answers, and the others tended not to falsify. However, differently than what would be expected in this case, the distance between the observed and theoretically expected falsification is not equally distributed across the whole range of a_p. Instead, the distance is significantly larger (observed falsification is lower than

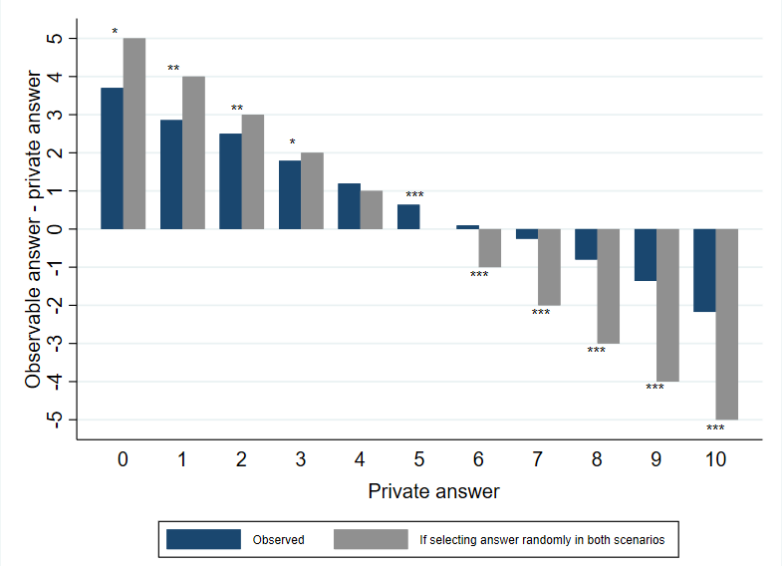

Figure 3. Observed and theoretical preference falsification

The figure depicts the average preference falsification, captured as a difference between the answers provided in the "observable" and in the "private" scenario ($\Delta_o = a_o - a_p$). The blue bars depict the average preference falsification observed in the sample. The grey bars depict preference falsification that would be expected if both a_p and a_o were selected randomly. Both values are depicted per answer provided in the private scenario. Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ in sign test for the equality of medians of observed (Δ_o) and the one that would result under random selection of a_o and a_p.

predicted) among those participants who privately indicated supportive attitudes ($a_p > 5$), than among those who indicated critical attitudes ($a_p < 5$). Additionally, Figure 4 depicts the share of participants who falsified upwards ($\Delta_o > 0$) in the upper panel, and the share of those who falsified downwards ($\Delta_o < 0$) in the lower panel, over a_p. As evident from the figure, the observed probability of falsification in both directions discontinuously changes around the neutral position indicated privately ($a_p = 5$), which would not be observed in the case of participants randomly selecting a_p and a_o. In particular running a probit regression of a dummy variable for observing positive (respectively negative) preference falsification, i.e., $\Delta_o > 0$ (resp. $\Delta_o < 0$) on a_p and a dummy variable ϕ that takes value one if $a_p < 5$ (resp. $a_p > 5$), yields a positive and significant coefficient for ϕ (with $p = 0.03$ and $p = 0.007$ respectively). Taken together, these findings suggest that the preference falsification was rather driven by the perceived social appropriateness of expressed views than by a random behavior.

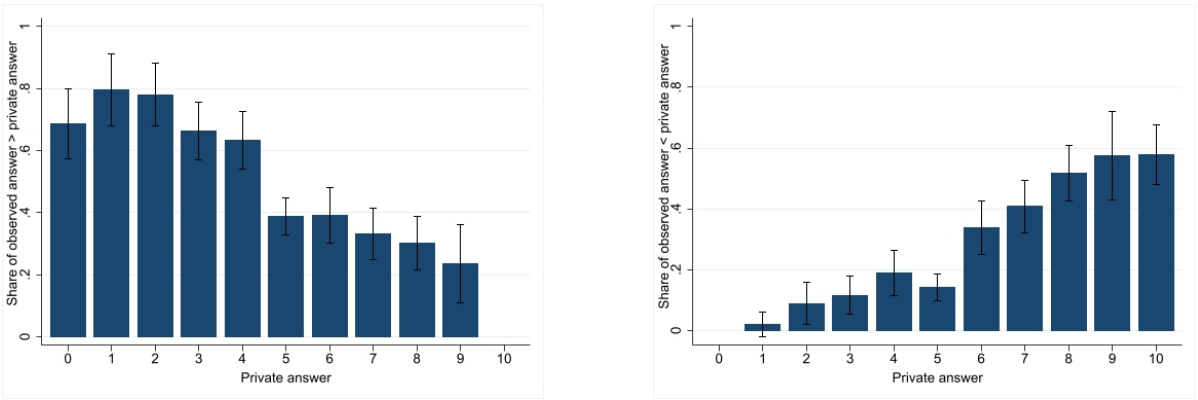

Figure 4. Share of participants with $\Delta_o > 0$ (left) and participants with $\Delta_o < 0$ (right)

The upper panel depicts the share of participants who provided a higher answer (a more positive view) in "observable" than in the "private" scenario ($a_o > a_p$). The lower panel depicts the share of participants who provided a lower (a more critical view) answer in "observable" than in the "private" scenario ($a_o < a_p$). Vertical lines designate the 95% confidence intervals.

Table A1. Sample description: Regions of origin

	Positive treatment	Negative treatment	Total
Region of (parental) origin			
Bulgaria & Romania	0.065	0.078	0.072
Central-Eastern European Union (Czech Republic, Slovakia, Poland, Hungary)	0.182	0.172	0.177
Baltic states (Estonia, Lithuania, Latvia)	0.011	0.007	0.009
Ex-Yugoslavia (Bosnia and Herzegovina, Croatia, Kosovo, Montenegro, North Macedonia, Serbia, Slovenia)	0.058	0.084	0.072
North Africa (Morocco, Algeria, Lybia, Tunesia and Egypt)	0.078	0.068	0.073
Southern European Union countries (Greece, Italy, Portugal, Spain, Cyprus and Malta)	0.106	0.145	0.127
Turkey	0.249	0.205	0.226
Southern Ex-Soviet union (Tajikistan, Turkmenistan, Georgia, Kazakhstan, Kyrgyzstan, Armenia and Azerbaijan)	0.063	0.073	0.068
Western Ex-Soviet union (Ukraine, Moldova, Belarus)	0.033	0.028	0.030
Russian federation	0.092	0.084	0.088
Albania	0.063	0.056	0.060
Age			
[18-24]	0.338	0.349	0.343
[25-34]	0.300	0.284	0.292
[35-44]	0.182	0.175	0.179
[45-54]	0.108	0.116	0.112
[55-64]	0.060	0.063	0.061
[65-74]	0.011	0.010	0.010
[75-84]	0.002	0.003	0.003
Gender			
Male	0.457	0.438	0.447
Education			
Primary or lower secondary	0.354	0.331	0.342
Secondary	0.233	0.238	0.236
Tertiary	0.413	0.431	0.423
Observations	554	605	1,159

Notes: Regions of participants' own or parental origin across treatments.

Table A2. Treatment effects: Pledged donation to the UNHCR

	(1)	(2)
	Pledged donation	
Negative treatment	-4.383***	-4.331***
	(1.308)	(1.301)
Age		-0.186**
		(0.074)
Tertiary education		1.417
		(1.781)
Equivalent household income		0.000
		(0.001)
Gender (1 if male)		0.728
		(1.380)
Constant	45.801***	50.182***
	(1.794)	(2.362)
Observations	1,159	1,159
R-squared	0.040	0.044

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Column (1) and column (2) depict the results of the OLS regression of the amount that participants pledged to donate to the UNHCR on the treatment variable and the set of individual controls. Negative treatment indicates receiving negative status information regarding own in-group (with Positive treatment serving as a baseline). All regressions include fixed effects of the federal state of residence in Germany and the region of participants' (parental) origin. Standard errors in parentheses are clustered on the level of region of participants' (parental) origin.

Table A3. Treatment effects: role of participants' socio-economic characteristics

	(1)	(2)
	Pledged donation	
Negative treatment	-4.678**	-6.398***
	(2.001)	(2.122)
Negative treatment*Income	-2.536	
	(2.554)	
Negative treatment*Tertiary Education		-2.175
		(6.091)
Income	2.529	1.216
	(2.898)	(2.024)
Tertiary Education	1.347	2.606
	(3.202)	(5.399)
Constant	53.595***	54.603***
	(4.480)	(4.381)
Individual controls	No	Yes
Observations	1,159	1,159

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Column (1) shows Tobit regression of the amount pledged to donate to the UNHCR on the treatment variable, equivalent household income tertile, and their interaction. Column (2) shows the results of the Tobit regression of the amount pledged to donate to the UNHCR on the treatment variable, indicator variable for tertiary education, and their interaction. All regressions include fixed effects of the federal state of residence in Germany and the region of participants' (parental) region of origin. Standard errors are clustered on the level of participants' (parental) region of origin.

References

Alesina, Alberto and Marco Tabellini, "The political effects of immigration: Culture or economics?," Technical Report, National Bureau of Economic Research 2022.

_ , Armando Miano, and Stefanie Stantcheva, "Immigration and redistribution," *The Review of Economic Studies*, 2023, *90* (1), 1–39.

Alexander, Richard D, *The biology of moral systems*, Transaction Publishers, 1987.

Álvarez-Benjumea, Amalia and Fabian Winter, "The breakdown of antiracist norms: A natural experiment on hate speech after terrorist attacks," *Proceedings of the National Academy of Sciences*, 2020, *117* (37), 22800–22804.

Bardsley, Nicholas and Rupert Sausgruber, "Conformity and reciprocity in public good provision," *Journal of Economic Psychology*, 2005, *26* (5), 664–681.

Barr, Abigail, Tom Lane, and Daniele Nosenzo, "On the social inappropriateness of discrimination," *Journal of Public Economics*, 2018, *164*, 153–164.

Benjumea, Amalia Álvarez, "Uncovering hidden opinions: social norms and the expression of xenophobic attitudes," *European Sociological Review*, 2022.

Berry, John W, "Mutual attitudes among immigrants and ethnocultural groups in Canada," *International journal of intercultural relations*, 2006, *30* (6), 719–734.

Bicchieri, Cristina and Erte Xiao, "Do the right thing: but only if others do so," *Journal of Behavioral Decision Making*, 2009, *22* (2), 191–208.

_ , Eugen Dimant, and Silvia Sonderegger, "It's not a lie if you believe the norm does not apply: conditional norm-following with strategic beliefs," *Available at SSRN 3326146*, 2020.

_ , _ , Simon Gaechter, and Daniele Nosenzo, "Observability, Social Proximity, and the Erosion of Norm Compliance," Technical Report 2020.

Bradley, Margaret M and Peter J Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of behavior therapy and experimental psychiatry*, 1994, *25* (1), 49–59.

Bursztyn, Leonardo, Alessandra L González, and David Yanagizawa-Drott, "Misperceived social norms: Female labor force participation in Saudi Arabia," Technical Report, National Bureau of Economic Research 2018.

_ , Georgy Egorov, and Stefano Fiorin, "From extreme to mainstream: The erosion of social norms," *American economic review*, 2020, *110* (11), 3522–48.

Card, David, Christian Dustmann, and Ian Preston, "Immigration, wages, and compositional amenities," *Journal of the European Economic Association*, 2012, *10* (1), 78–119.

Crandall, Christian S, Amy Eshleman, and Laurie O'brien, "Social norms and the expression and suppression of prejudice: the struggle for internalization.," *Journal of personality and social psychology*, 2002, *82* (3), 359.

Dandy, Justine and Rogelia Pe-Pua, "Attitudes to multiculturalism, immigration and cultural diversity: Comparison of dominant and non-dominant groups in three Australian states," *International Journal of Intercultural Relations*, 2010, *34* (1), 34–46.

DellaVigna, Stefano, John A List, Ulrike Malmendier, and Gautam Rao, "Voting to tell others," *The Review of Economic Studies*, 2016, *84* (1), 143–181.

Derks, Belle, Colette Van Laar, and Naomi Ellemers, "The queen bee phenomenon: Why women leaders distance themselves from junior women," *The Leadership Quarterly*, 2016, *27* (3), 456–469.

Dinas, Elias, Sergi Martínez, and Vicente Valentim, "Social Norm Change, Political Symbols, and Expression of Stigmatized Preferences," *Political Symbols, and Expression of Stigmatized Preferences (January 11, 2020)*, 2020.

_ , Vasiliki Fouka, and Alain Schläpfer, "Family history and attitudes toward outgroups: evidence from the European refugee crisis," *The journal of politics*, 2021, *83* (2), 647–661.

Dunin-Wasowicz, Roch, "Why did South Asians vote for Brexit?," *LSE Blog*, 03 November 2016. Available: https://blogs.lse.ac.uk/brexit/2016/11/03/why-did-south-asians-vote-for-brexit/ [Last accessed: 17 December 2023] 2016.

_ , "The British Asian vote for Brexit contains a few surprises," *LSE Blog*, 20 February 2017. Available: https://blogs.lse.ac.uk/brexit/2017/02/20/the-british-asian-vote-for-brexit-contains-a-few-surprises/ [Last accessed: 17 December 2023] 2017.

Ellemers, Naomi, Henriette Van den Heuvel, Dick De Gilder, Anne Maass, and Alessandra Bonvini, "The underrepresentation of women in science: Differential commitment or the queen bee syndrome?," *British Journal of Social Psychology*, 2004, *43* (3), 315–338.

Enikolopov, Ruben, Alexey Makarin, Maria Petrova, and Leonid Polishchuk, "Social image, networks, and protest participation," *Networks, and Protest Participation (April 26, 2020)*, 2020.

Facchini, Giovanni and Anna Maria Mayda, "Does the welfare state affect individual attitudes toward immigrants? Evidence across countries," *The review of economics and statistics*, 2009, *91* (2), 295–314.

Faniko, Klea, Naomi Ellemers, Belle Derks, and Fabio Lorenzi-Cioldi, "Nothing changes, really: Why women who break through the glass ceiling end up reinforcing it," *Personality and Social Psychology Bulletin*, 2017, *43* (5), 638–651.

Fouka, Vasiliki, Soumyajit Mazumder, and Marco Tabellini, "Changing In-Group Boundaries: The Effect of Immigration on Race Relations in the US," 2020.

_ , _ , **and** _ , "From Immigrants to Americans: Race and Assimilation during the Great Migration," *The Review of Economic Studies*, 06 2021. rdab038.

Froehlich, Laura and Isabel Schulte, "Warmth and competence stereotypes about immigrant groups in Germany," *Plos one*, 2019, *14* (9), e0223103.

Gerber, Alan S, Donald P Green, and Christopher W Larimer, "Social pressure and voter turnout: Evidence from a large-scale field experiment," *American political Science review*, 2008, *102* (1), 33–48.

Giuliano, Paola, Marco Tabellini et al., "The seeds of ideology: Historical immigration and political preferences in the United States," Technical Report, National Bureau of Economic Research 2020.

Goerres, Achim, Dennis C. Spies, and Sabrina Mayer, "Immigrant German Election Study (IMGES)," GESIS Datenarchiv, Köln. ZA7495 Datenfile Version 1.0.1, https://doi.org/10.4232/1.13544 2020.

Greiner, Ben and M Vittoria Levati, "Indirect reciprocity in cyclical networks: An experimental study," *Journal of Economic Psychology*, 2005, *26* (5), 711–731.

Haaland, Ingar and Christopher Roth, "Labor market concerns and support for immigration," *Journal of Public Economics*, 2020, *191*, 104256.

Hainmueller, Jens and Daniel J. Hopkins, "Public Attitudes Toward Immigration," *Annual Review of Political Science*, 2014, *17* (1), 225–249.

_ **and Dominik Hangartner**, "Who gets a Swiss passport? A natural experiment in immigrant discrimination," *American political science review*, 2013, *107* (1), 159–187.

_ **, Michael J Hiscox, and Yotam Margalit**, "Do concerns about labor market competition shape attitudes toward immigration? New evidence," *Journal of International Economics*, 2015, *97* (1), 193–207.

Hindriks, Paul, Maykel Verkuyten, and Marcel Coenders, "Evaluating political acculturation strategies: The perspective of the majority and other minority groups," *Political Psychology*, 2017, *38* (5), 741–756.

Hugo, Rich Morin Lopez Mark and Paul Taylor, "Illegal Immigration Backlash Worries, Divides Latinos," *Pew research Center*, 2010.

Kashima, Yoshihisa, Samuel Wilson, Dean Lusher, Leonie J Pearson, and Craig Pearson, "The acquisition of perceived descriptive norms as social category learning in social networks," *Social Networks*, 2013, *35* (4), 711–719.

Krupka, Erin and Roberto A Weber, "The focusing and informational effects of norms on pro-social behavior," *Journal of Economic psychology*, 2009, *30* (3), 307–320.

Kuran, Timur, *Private truths, public lies: The social consequences of preference falsification*, Harvard University Press, 1997.

Kwan, Letty Y-Y, Suhui Yap, and Chi yue Chiu, "Mere exposure affects perceived descriptive norms: Implications for personal preferences and trust," *Organizational Behavior and Human Decision Processes*, 2015, *129*, 48–58.

Lange, Jason, Bo Erickson, and Brad Heath, "Trump's Return to Power Fueled by Hispanic Working-Class Voter Support," November 2024. Accessed: 2024-12-16.

Mayda, Anna Maria, "Who is against immigration? A cross-country investigation of individual attitudes toward immigrants," *The review of Economics and Statistics*, 2006, *88* (3), 510–530.

Meeusen, Cecil, Koen Abts, and Bart Meuleman, "Between solidarity and competitive threat?: The ambivalence of anti-immigrant attitudes among ethnic minorities," *International Journal of Intercultural Relations*, 2019, *71*, 1–13.

Meidert, Nadine and Carolin Rapp, "Public Attitudes towards Refugees in Germany: What Drives Attitudes towards Refugees in Comparison with Immigrant Workers from European Union Countries?," *Journal of Refugee Studies*, 12 2019, *32* ($\text{Special}_I ssue_1$), $i209 - -i218$.

Moriconi, Simone, Giovanni Peri, and Riccardo Turati, "Are Immigrants more Left leaning than Natives?," Technical Report, National Bureau of Economic Research 2022.

Mujcic, Redzo and Andreas Leibbrandt, "Indirect reciprocity and prosocial behaviour: evidence from a natural field experiment," *The Economic Journal*, 2018, *128* (611), 1683–1699.

Nowak, Martin A and Karl Sigmund, "Evolution of indirect reciprocity," *Nature*, 2005, *437* (7063), 1291–1298.

Pardos-Prado, Sergi and Carla Xena, "Skill specificity and attitudes toward immigration," *American Journal of Political Science*, 2019, *63* (2), 286–304.

Perez-Truglia, Ricardo and Guillermo Cruces, "Partisan interactions: Evidence from a field experiment in the united states," *Journal of Political Economy*, 2017, *125* (4), 1208–1243.

Pettersson, Katarina, Karmela Liebkind, and Inari Sakki, "You who are an immigrant–why are you in the Sweden Democrats?," *Discourse & Society*, 2016, *27* (6), 624–641.

Ridgeway, Cecilia L, *Status: Why is it everywhere? Why does it matter?*, Russell Sage Foundation, 2019.

Sahgal, Neha, Alan Cooperman, and A Schiller, "Being Christian in Western Europe," *Pew research Center*, 2018.

Scheve, Kenneth F and Matthew J Slaughter, "Labor market competition and individual preferences over immigration policy," *Review of Economics and Statistics*, 2001, *83* (1), 133–145.

Shackle, Samira, "Meet the immigrants who voted for Brexit," *Politico Blog*, 25 August 2016. Available: https://www.politico.eu/article/immigrants-who-voted-for-brexit-luton-migration/ [Last accessed: 17 December 2023] 2017.

Strijbis, Oliver and Javier Polavieja, "Immigrants against immigration: Competition, identity and immigrants' vote on free movement in Switzerland," *Electoral Studies*, 2018, *56*, 150–157.

Tabellini, Marco, "Gifts of the immigrants, woes of the natives: Lessons from the age of mass migration," *The Review of Economic Studies*, 2020, *87* (1), 454–486.

Tajfel, Henri, "Cognitive aspects of prejudice," *Journal of biosocial science*, 1969, *1* (S1), 173–191.

_ , John C Turner, William G Austin, and Stephen Worchel, "An integrative theory of intergroup conflict," *Organizational identity: A reader*, 1979, *56*, 65.

Turner, John C, "Social comparison and social identity: Some prospects for intergroup behaviour," *European journal of social psychology*, 1975, *5* (1), 1–34.

Valentim, Vicente, "Parliamentary representation and the normalization of radical right support," *Comparative Political Studies*, 2021, *54* (14), 2475–2511.

_ , "Political Stigmatization and Preference Falsification: Theory and Observational Evidence," *Available at SSRN 4023263*, 2022.

van Bezouw, Maarten Johannes, Jojanneke van Der Toorn, and Julia Christina Becker, "Social creativity: Reviving a social identity approach to social stability," *European Journal of Social Psychology*, 2021, *51* (2), 409–422.

Weber, Max, "Economy and Society, edited by G. Roth and C. Wittich," *New York: Bedminster*, 1968.