



ARTICLE



<https://doi.org/10.1057/s41599-025-05549-6>

OPEN

# Does anti-immigrant sentiment decrease support for redistribution? Evidence from two online experiments

Alice **Guerra** <sup>1✉</sup> & Brooke **Harrington**<sup>2</sup>

This study examines a relationship that has been implied by years of correlational research: that natives' support for welfare redistribution declines when benefits are allocated to immigrants rather than natives—a phenomenon known as welfare chauvinism. We conducted two online experiments ( $N_1 = 273$ ,  $N_2 = 1060$ ) involving redistribution to unemployed people through real donations to charity, framed as tax compliance decisions within a simulated reporting task. We employed a between-subject design, randomly assigning participants to treatments that differed solely in the immigration status of the charity-benefit recipients. Drawing from native samples in Italy, Denmark, and the UK, we find that natives' support for redistribution is not statistically affected by the recipients' immigration status. This null effect holds across both studies, despite spanning a four-year period (2000–2024) marked by major global events which might have been expected to shift preferences regarding welfare state distribution and immigration, including: the COVID-19 pandemic, the Ukrainian refugee crisis, and the increasingly anti-immigrant turn in Italian, Danish, and UK politics. Our findings challenge prevailing theories of welfare chauvinism and invite both replication efforts and reconsideration of long-standing theoretical givens.

UNCORRECTED PROOF

<sup>1</sup>Department of Economics, University of Bologna, Bologna, Italy. <sup>2</sup>Department of Sociology, Dartmouth College, Hanover, NH, USA.  
✉email: [alice.guerra3@unibo.it](mailto:alice.guerra3@unibo.it)

## Introduction

Individuals' support for welfare redistribution is contingent on their beliefs about the "deservingness" of the beneficiaries (e.g., Reskins and van der Meer, 2017; Borisova et al. 2018). Prior correlational studies have shown that people express the strongest support for welfare redistribution when recipients are similar to themselves on key demographic characteristics, such as native-born status and race (Gilens, 2000; Luttmer, 2001; Lago-Peñas and Lago-Peñas, 2010; Fong and Luttmer, 2011; Reeskens and Van der Meer, 2019; Almås et al. 2020; Alesina et al. 2023). Employment status is also a key vector of deservingness, with employed people viewed as more deserving than the unemployed (Kootstra, 2017; Nicholls, 2020).

The close connection between group identities and support for redistribution is known as "welfare chauvinism" (Van der Waal et al. 2010; Dahlberg et al. 2012; Reeskens and Van der Meer, 2019). The flip side of the phenomenon is that social heterogeneity decreases support for welfare redistribution. That is, when individuals perceive that benefits are going to those different from themselves, their support for redistribution declines (Alesina and Glaeser, 2004; Cullen et al. 2021; Alesina et al. 2023).

In many parts of the world, immigration status is among the most salient forms of interpersonal difference, with immigrants often perceived as the *least* deserving category of all—even less deserving of welfare benefits than unemployed natives (van Oorschot, 2006; Dahlberg et al. 2012; Tabellini, 2020; Fouka and Tabellini, 2022).<sup>1</sup> Indeed, immigrants who receive benefits from welfare states are frequently regarded by natives as the "undeserving poor" compared to other vulnerable groups, such as native elderly and unemployed (Bommes and Geddes, 2000; Reeskens and Van der Meer, 2019).

Correlational studies have also found that individuals' perception of immigrants as fundamentally outside the national polity and therefore undeserving of welfare benefits is negatively correlated with their intrinsic motivation to pay taxes—what is often termed their "tax morale" (e.g., Finseraas, 2008; Rodriguez-Justicia and Theilen, 2022; for reviews, see Alm, 2019; Guerra and Harrington, 2021; Van der Meer and Reeskens, 2021). For example, Nemore and Morone (2019) analyzed survey data from the European Values Survey and showed that natives with anti-immigrant attitudes also have lower tax morale. By relying upon a large-scale survey conducted across six countries (Germany, France, Italy, Sweden, the UK, US), Alesina et al. (2023) found that the perception of immigrants as economically weaker, hence more likely to rely on the welfare system, is strongly correlated with lower support for redistribution, as captured by individuals' preference for a less progressive income tax system, less redistributive spending, and reduced real donations to charity.

In sum, prior research suggests that natives' negative attitudes toward immigrants are correlated with reduced support for redistribution (van Oorschot, 2006; Eger, 2010; Bauböck and Scholten, 2016; Buss, 2019). While this relationship has been well-documented in correlational studies, to our knowledge the causal link has not yet been cleanly tested. As a preliminary effort to address this gap, we conducted two online experiments involving redistribution to unemployed people through (real) donations to charity, framed as tax compliance decisions in a simulated reporting task. We employed a between-subject design, randomly assigning participants to treatments that differed solely in the immigration status of the donation beneficiaries, who were either natives or immigrants (no reference to the immigration status in the control group).

More specifically, relying on well-established experimental designs from the tax evasion literature (for a comprehensive review, see Alm and Malézieux, 2021), participants first completed a real-effort clerical task—adapted from Zhang et al. (2016)

and Pampel et al. (2019)—and were then asked to decide how much of their earned income to report, knowing that reported amounts would be taxed at a fixed rate and could be audited with monetary penalties for underreporting.

Building on prior research using tax evasion experiments with redistribution to real-world public goods (e.g., to the Red Cross, as in Doerrenberg and Duncan, 2014, and Doerrenberg, 2015), the simulated "tax revenues" collected in our study were donated to the Red Cross, which redistributed the donations to our targeted beneficiaries.<sup>2</sup> As pointed out by Doerrenberg and Duncan (2014: 52), "donating tax revenues to the Red Cross may be viewed as [a] public good, which has been shown to have a positive effect on willingness to pay taxes." The use of tax compliance games involving real monetary redistribution to charities or public goods is not new and has grown substantially in the last decade. For example, Doerrenberg (2015) used a tax evasion experiment to compare compliance rates in response to varying uses of tax revenues, including: redistribution to participants, donations to the Red Cross, funding of the experimenter's research, and transfers to the federal government. They found no significant differences in "tax" compliance across these conditions. Likewise, Alm and Malézieux's (2021) meta-analysis shows that "tax"-funded donations to real-life public goods—including charities—produce "tax" compliance effects comparable to those achieved via participant redistribution.

Aligned with this tradition in experimental research design, we used donations framed to simulate tax compliance as a proxy for redistribution. We then manipulated exogenously the immigration status of the redistribution beneficiaries. Specifically, participants were randomly assigned to one of three treatment conditions: redistribution to *unemployed natives*, redistribution to *unemployed immigrants*, or redistribution to *unemployed individuals* with no reference to immigration status (control group). The three treatments were identical in all respects except for the immigration status of the donation recipients. Therefore, any differences in participants' support for redistribution through "tax" compliance can be attributed *solely* to this treatment variation, i.e., only to the immigration status of the redistribution beneficiaries.

While the experimental method has inherent limitations in terms of external validity, it is particularly well-suited to our research question, which would be extremely difficult to address using observational data, as is nearly impossible to isolate recipients' identity as a causal factor in real-world redistribution contexts. This is the advantage of the experimental approach, achieved through randomly assigning participants to treatment groups while keeping all other variables constant. This methodology enabled us to conduct identical experiments in three national contexts—Italy, Denmark, and the UK—that differ significantly in terms of compliance norms and political discourse around immigration and welfare (Reeskens and Van Oorschot, 2012; Van Der Waal et al. 2013; Jørgensen and Thomsen, 2016; Kim and Lee, 2021). Moreover, the redistribution mechanism (through donations to the Red Cross) was held constant across all conditions: only the immigration status of the donation recipients varied. As such, any differences in participants' reporting behavior can be attributed solely to the randomized immigration status of the beneficiaries, not to the donation structure itself.

Our results run counter to the expectations established by previous correlational and observational studies. Rather than finding that support for redistribution (through "tax" compliance) is affected by welfare chauvinism, we show that the immigration status of the beneficiaries has no statistically significant effect on the reporting behavior of natives. These findings also contribute to the economic literature on "in-group"

Q1–Q5

favoritism, i.e., the tendency to favor or evaluate more positively one's own social group over the others (Carrarini and Mengel, 2016; Lane, 2016). With few exceptions (e.g., Fershtman and Gneezy 2001, Berge et al. 2020), evidence from lab and lab-in-the-field experiments have consistently shown that people are more altruistic and trusting towards members of their own social groups, whether the groupings are based on blood or ethnic ties, nationality, age or race, etc. (e.g., Bernhard et al. 2006; Goette et al. 2006; Enke et al. 2022). The evidence has been more mixed in studies looking at the flip side of in-group favoritism—discrimination against out-groups (e.g. Lane, 2016; Felfe et al. 2021). Importantly, “out-group” discrimination tends to be stronger in experiments using artificially-induced group identities than in those using natural social identities (Barr et al. 2018). Our findings add evidence to this literature, supporting the conclusion that “out-group” discrimination based on natural identity (here, immigration status) is less pronounced than is theoretically expected.

In evaluating our findings, as with those of any experimental simulation, some caution is warranted when extrapolating results to real-world behavior. We encourage replication efforts and further research, especially given that our findings conflict with traditional conceptions of welfare chauvinism and call into question long-standing theoretical assumptions.

The remainder of the paper is organized as follows. Section 2 outlines the experimental design and methodology that are common to our two studies. Section 3 presents the procedures and results from the Pilot Study ( $N = 273$ ), an online experiment conducted in collaboration with the Laboratory for Research in Behavioral Experimental Economics (LINEEX) of the University of Valencia in Spain. Section 4 details the procedures and results of the Main Study ( $N = 1060$ ), which extends our pilot analysis through an online experiment on the Prolific platform. Section 5 discusses key insights, limitations, and avenues for future research. The Online Supplementary Material includes additional tables and figures, balance tables, and detailed experimental instructions.

## Experimental design and methodology

**Experimental design.** Our experiments involve real donations to charity, framed to simulate redistribution funded by tax compliance. Our research design follows a well-established framework from the experimental literature on tax compliance: the one-shot tax evasion game (for a review, see Alm and Malézieux, 2021).<sup>3</sup> In keeping with that tradition, participants were asked to perform a clerical task, for which they earned points (converted into their national currency)—earnings that were then “taxed” based on participants’ reporting choices. The “taxes” collected, along with fines when participants were audited and found to have under-reported their earnings, created the pool of funding donated to charity. All of this was explained in advance of the game to participants, framed in a way that explicitly and repeatedly emphasized the “tax” compliance (earnings reporting) decisions they would be asked to make, under “tax” rates and audit probabilities we defined, and with specific fines for under-reporting. Throughout this section, we use the same fiscal terminology presented to participants in our experiments—e.g., by referring to “taxes,” “reporting,” and “audits.” We stress that these terms reflect the framing of a *simulated* reporting task and do not imply real-world tax liability.

The game consists of three parts, as summarized in Fig. 1: (1) the real-effort task, where participants had the opportunity to gain points that were converted into their national currency at the end of the experiment;<sup>4</sup> (2) the reporting task, where participants were asked to report their points for redistribution purposes; and

(3) the post-experiment questionnaire, where participants were asked questions about their socio-demographic characteristics, attitudes towards immigration, and risk preferences.<sup>5</sup>

Specifically, the real-effort task was the same for all participants and consisted of counting the number of “a”s in a paragraph, within a two-minute time limit. We chose this type of task since it does not require any prior knowledge or skills and involves no learning strategy. At the beginning of the experiment, participants were informed about the payment structure: that they would earn a maximum of 100 points if they estimated the exact number of “a”s correctly; otherwise, their earnings decreased from 100 to 0 according to the difference between the provided answer and the correct value.<sup>6</sup>

Once the two minutes for the task ended, the reporting phase began.<sup>7</sup> Participants were asked to report their points for redistribution purposes. They were informed that they were free to report any amount from zero to 100% of their actual points earned, and that only the points they reported would be “taxed” at the 30% rate—all subject to a 5% probability of being audited. If audited and caught under-reporting their earned points, participants had to pay a fine equal to twice the amount of “taxes” they evaded. Participants were informed that neither their reporting decision nor the potential audits would affect their anonymity.

The “taxes” and the fines paid as a result of the audits constitute the pool of funding donated to charity (the Red Cross) at the end of all the experimental sessions.<sup>8</sup> We exogenously manipulated the immigration status of the donation recipients through a between-subject design with three treatments. In the “Control” treatment, we told participants that a charity would redistribute their “tax” contributions as donations to *unemployed people*. In the “Natives” treatment, the charity redistributed participants’ “tax” contributions as donations to *unemployed natives*—those born in Italy, Denmark, or the UK. In the “Immigrants” treatment, the charity redistributed participants’ “tax” contributions as donations to *unemployed immigrants*—those not born in Italy, Denmark, or the UK.

From the beginning of the experiment, before starting the real-effort task, participants were informed about the redistribution scheme and were reminded about it throughout the experiment, including at the reporting stage. After the reporting task, they were asked a variety of background and attitudinal questions (with no consequences on their final earnings). After completing the questionnaire, participants received information about their final earnings and the payment procedure.<sup>9</sup>

**Measures.** Our analysis tests whether and to what extent the immigration status of the redistribution beneficiaries (here, charity-benefit recipients) affects individuals’ support for redistribution (here, participants’ willingness to report their true earnings in the redistribution simulation task). We estimate this effect using linear regression models, where the main outcome variable is compliance, measured as the ratio of reported points to actually earned points in the real-effort task. Potential values range from 0 (full evasion) to 1 (full compliance). As it is common in many experiments on tax compliance, this variable has a multimodal distribution, with spikes at 0 and 1, in both the pooled sample and the country-specific samples (see Supplementary Appendix A Figs S1 and S2 online).

Our key independent variables are treatment dummies:

- “Control Treat.” (baseline), wherein the Red Cross redistributed participants’ “tax” contributions as donations to *unemployed people*—including both immigrants and natives (those born in Italy, Denmark, and the UK);
- “Natives Treat.,” wherein the Red Cross redistributed participants’ “tax” contributions as donations to

Real-effort task	→	Reporting Task	→	Questionnaire
<ul style="list-style-type: none"> <li>Counting the number of “a”s in the text in a paragraph.</li> <li>Two-minute time limit.</li> <li>Earn a maximum of 100 points if they estimated the exact number of “a”s correctly.</li> <li>Otherwise, earnings decrease from 100 to zero according to the difference between the provided answer and the correct value.</li> </ul>		<ul style="list-style-type: none"> <li>Free to report any amount from zero to 100% of the actual points earned.</li> <li>Only the points reported are taxed at the 30% tax rate.</li> <li>5% probability of being audited.</li> <li>If audited and caught under-reporting the actual points, pay a fine equal to twice the amount of “taxes” evaded.</li> <li>The “taxes” and the fines paid as a result of the audits constitute the pool of funding donated to charity at the end of all the experimental sessions. See the “Redistribution Stage.”</li> </ul>		Set of background and attitudinal questions (with no consequences on final earnings).
<b>Redistribution Stage</b>				
<b>Between-subject design with three treatments, corresponding to three different recipient pools</b>				
“Control” treatment		The charity redistributes the “tax” contributions as donations to <b>unemployed people</b> .		
“Natives” treatment		The charity redistributes the “tax” contributions as donations to <b>unemployed natives</b> —those born in Italy, Denmark, or the UK.		
“Immigrants” treatment		The charity redistributes the “tax” contributions as donations to <b>unemployed immigrants</b> —those not born in Italy, Denmark, or the UK.		

Fig. 1 Design of the experiment.

*unemployed natives*—i.e., only those born in Italy, Denmark, or the UK;

- “Immigrants Treat.,” wherein the Red Cross redistributed participants’ “tax” contributions as donations to *unemployed immigrants*—i.e., only those *not* born in Italy, Denmark, or the UK.

Here, we are interested in analyzing the treatment effects on average compliance across countries. We do not compare average compliance between countries; this analysis has already been conducted by Guerra and Harrington (2018) using a similar experimental design but a different sample and method, as well as by Pampel et al. (2019).<sup>10</sup>

At the end of the survey, we gathered data on individual characteristics shown in previous research to shape participants’ compliance (Torgler, 2006; D’Attoma et al. 2017; Ottone et al. 2018; Nemore and Morone, 2019; Guerra and Harrington, 2021, 2023). These characteristics include: gender (1 for males); age; dummies for each education level; dummies for each type of employment status; having Economics as an educational background; approximate monthly household income, after taxes and other deductions; and having participated in other similar experimental studies. We also collected information on individuals’ political stance (1 for right; 0 otherwise),<sup>11</sup> as well as their membership in or contact with any humanitarian or charitable organizations, their frequency of attending religious services (apart from weddings, funerals, and christenings), their risk preferences (1 for risk-avoiding, 0 for risk-seeking), and their attitudes towards immigrants. To measure attitudes toward immigrants, we drew on the following items from the European Value Survey (EVS) and asked participants to what extent they agreed:

- Immigrants take jobs away from natives.
- Immigrants make crime problems worse.

- Immigrants are a strain on the Italian/Danish/UK welfare system.
- Today in Italy/Denmark/the UK, there are too many immigrants.

As in the original EVS, responses to each of those statements were coded into a 5-point scale, in which 0 stands for ‘Strongly disagree,’ 1 for ‘Disagree,’ 2 for ‘Neither agree nor disagree,’ 3 for ‘Agree,’ and 4 for ‘Strongly agree.’<sup>12</sup>

**Pilot study: the LINEEX experiment**

**Procedures.** Our Pilot Study was programmed using oTree (Chen et al. 2016), and conducted online in November 2020 in Italy and Denmark. It followed the standards and procedures that routinely guide conventional experiments in economics (e.g., anonymity, no deception, and real payments).

In the interest of time and technical efficiency, in each country we ran the three treatments at the same time. That is, we conducted one session in each country, randomly allocating participants to the three treatments as they entered the experiment. This resulted in an uneven distribution of observations across treatments, as shown in Table 1.

We devoted particular care to recruiting our sample, in collaboration with the Laboratory for Research in Behavioral Experimental Economics (LINEEX) of the University of Valencia in Spain. We contracted with LINEEX for samples of citizens aged at least 18 years of age in both Italy and Denmark. A custom marketing campaign on social networks—including Facebook, Instagram, and YouTube—was built to announce the experiment. That campaign was targeted to reach a broad segment of the population, including both students and non-students; it ran for about eight weeks in total. This marketing campaign allowed us to build a database of 600 potential participants—300 in Denmark and 300 in Italy.<sup>13</sup> Next, to construct a sample suitable to address

**Table 1 Number of observations, by treatment and country (pilot study).**

Treatment	Italy	Denmark	Pooled sample
Control Treat.	41 30.37%	37 26.81%	78 28.57%
Natives Treat.	50 37.04%	52 37.68%	102 37.36%
Immigrants Treat.	44 32.59%	49 35.51%	93 34.07%
Total	135	138	273

**Table 2 Participants' demographics and other characteristics (Pilot Study).**

Variable	Mean	SD	Median	Min	Max
Italy	0.495	0.501	0	0	1
Sex (1 male)	0.582	0.494	1	0	1
Age	24.645	5.244	24	18	59
Student (0/1)	0.527	0.500	1	0	1
Economist (0/1)	0.315	0.465	0	0	1
Have kids (0/1)	0.048	0.213	0	0	1
Participated in other similar studies (0/1)	0.410	0.493	0	0	1
- If yes, how many	4.152	2.190	4.5	1	7
Amount of failed comprehension questions	1.051	1.688	1	0	13
<i>Education level</i>	Freq.	%	Cum.%		
Junior high school	4	1.465	1.465		
High school	103	37.729	39.194		
Bachelor Degree	85	31.136	70.330		
Master Degree	72	26.374	96.703		
Other	9	3.297	100		
<i>Monthly household income (in EUR)</i>	Freq.	%	Cum.%		
< EUR 3k	110	45.643	45.643		
EUR 3k to 6k	86	35.685	81.328		
EUR 6k to 10k	30	12.448	93.776		
> EUR 10k	15	6.224	100		
Total	241	100			
<i>Political stance</i>	Freq.	%	Cum.%		
Left or Center	196	71.795	71.795		
Right	77	28.205	100		
<i>Belong to and/or have contacts with any humanitarian or charitable organization</i>	Freq.	%	Cum.%		
No	174	63.736	63.736		
Yes	99	36.264	100		
<i>Frequency of attending religious services (apart from weddings, funerals, and christenings)</i>	Freq.	%	Cum.%		
More than once a week	2	0.733	0.733		
Once a week	14	5.128	5.861		
Once a month	13	4.762	10.623		
Only on specific holy days	73	26.740	37.363		
Once a year	23	8.425	45.788		
Practically never	148	54.212	100		
<i>How do you see yourself: Are you in general a person who takes risks or do you try to avoid risks?</i>	Freq.	%	Cum.%		
Take risk	139	50.916	50.916		
Avoid risk	134	49.084	100		
Total	273	100			

our research question and to ensure the accuracy of our data, we conducted a set of screening tests which led us to block a number of “unwanted” subjects from participating in our experiment. The first screen—which was directly administered to participants—was aimed at checking (i) their nationality and (ii) the nationality of their parents. To isolate welfare chauvinism, it was important

to ensure participants were natives—i.e., born in Italy/Denmark to Italian/Danish parents. This screen led us to block 192 of the original 600 subjects, of which 95 were in Italy and 97 in Denmark, because they failed the “being native” requirement.

Among the remaining 408 native participants, we conducted a second screen checking their GeoIP and Paypal accounts; this was done to identify where individuals were located and if the IP already participated in the experiment, as well as to spot fake or duplicate profiles. This led us to block another 51 subjects. Finally, one additional subject—a Dane—was excluded for not responding to the consent form presented to all participants at the beginning of the study.

Overall, these sets of screening tests ensured us excellent rates of data quality, though at the inevitable cost of having a narrower final sample than that initially planned. We ended up with 356 participants, all passing the checks and digitally signing the consent form. However, we lost another 83 individuals—45 Italians and 38 Danes—because they did not finish the experiment and/or did not collect the final payment. Hence, our final dataset consisted of 273 observations—of which 135 in Italy (49.45%) and 138 in Denmark (50.55%).

**Sample.** Table 2 reports the statistics of the pooled sample ( $N = 273$ ); see Supplementary Appendix A Table S1 online for additional descriptives.

About half of our sample were students (52.75%), with the other half either employed (35.90%) or retired, unemployed, or other status (11.35%). Overall, 58.24% of participants were male; the average age was 24.64 years old, with a range of 18 to 59 years. Similar proportions of the sample held high school diplomas (37.73%) and bachelor’s degrees (31.14%); another 26.37% had completed a master’s degree. Among all the participants, 31.50% indicated Economics as their main educational background (completed or in progress), followed by Arts and Humanities (11.36%) and Law (10.26%). Regarding other preferences, 28.21% of the sample described themselves as holding right-wing political views; 36.26% said they belonged to and/or had contacts with humanitarian or charitable organizations; and 10.62% said they attended religious services (other than weddings, funerals, and christenings) at least once a month. Attitudes towards risk were balanced, with 50.92% of subjects describing themselves as risk-takers, while the remaining 49.08% said they avoided risks. Beyond these core demographics, 41.0% of participants in the Pilot Study had previously participated in similar experiments, with a median of four prior participations. The mean number of failed comprehension questions was 1.051, indicating a generally good understanding of the experimental tasks.

Table 3 reports summary statistics about participants’ attitudes towards immigrants. Overall, participants generally reported low levels of anti-immigrant sentiment, particularly regarding economic concerns. A majority (84.25%) disagreed or strongly disagreed with the statement that immigrants take jobs away from natives; similarly, more than half (54.58%) disagreed or strongly disagreed with the claim that immigrants place a strain on the welfare system. Attitudes were slightly more divided when it came to crime and overall immigration levels. Still a majority (52.38%) disagreed that immigrants worsen crime problems, yet a significant portion of participants were neutral or even agreed (21.25% and 26.37%, respectively). Responses were more evenly distributed regarding whether there were “too many immigrants” in participants’ respective countries, with 55.68% disagreeing, 26.37% neutral, and 17.95% agreeing.

**Results.** Table 4 presents descriptive statistics on key experimental variables measured during the real-effort task and reporting stage.

Regarding the real-effort task, 39 out of 273 participants (14.29%) correctly counted the “a”s in the text in the paragraph, earning the maximum possible points. On average, participants earned 83.76 points, with a median of 88 points. Our key dependent variable is compliance in reporting earnings from the clerical task for “tax” purposes, measured as the ratio of earned points to reported points. Participants reported an average of 60.58 points to be “taxed,” resulting in an average compliance rate of 0.72. Eleven participants (4.03%) were randomly audited: eight truthfully reported their earned points, while three were caught underreporting their earned points and had to pay the fine.

Table 5 presents OLS regression estimates assessing treatment effects on compliance rates. Across all models, results show no statistically significant differences in compliance between treatment groups.

All specifications include treatment dummies. Models (2), (3), and (5) add the Italy dummy, and Model (3) further adds the

interactions between the Italy dummy and each treatment dummy. Models (4) and (5) add controls for subjects’ demographic characteristics, with Model (5) further including attitudinal controls. To test all treatment effects, Table 5 also reports the *p* values of Wald tests of hypotheses relating to the coefficients of Natives vs Immigrants and Natives×Italy vs Immigrants×Italy.

Overall, the coefficients for treatment groups remain small and not statistically significant across all specifications (*p* > 0.10). The inclusion of interaction terms (Model 3) and additional controls (Models 4 and 5) does not alter this conclusion.<sup>14</sup> The Wald tests confirm that the coefficients for Natives and Immigrants are not significantly different from each other (*p* > 0.10 in all models), further supporting the conclusion that treatment effects on compliance are negligible. These findings remain robust to alternative model specifications, including subsample analyses by country and gender (available upon request).

**Sample limitations and rationale for the main study.** Two features of our Pilot Study sample limit generalizability and statistical power. First, as in other online panel studies (e.g., Paolacci et al. 2010; Huff and Tingley, 2015), the sample is relatively young and politically left-leaning compared to the general population. Second, the null finding may to some extent be due to the relatively low number of observations.

To address these limitations, we conducted a second study with a larger and more representative sample: the average age of participants is significantly higher, and the majority are full-time employed (not students). Moreover, all participants are paying or have paid income tax in the real world. We also expanded geographic coverage to include the UK, in addition to Italy and Denmark.

From an experimental economics perspective, prior studies show that participants from Italy, Denmark, and the UK respond in broadly similar ways to tax-related incentives and redistribution schemes (Guerra and Harrington, 2018; Zhang et al. 2016; Pampel et al. 2019). While average compliance levels are higher among Italians, the behavioral responses to experimental manipulations—such as redistribution schemes—are remarkably consistent across national samples. This supports the validity of continuing, as in the Pilot Study, to pool the data throughout the analysis, allowing us to isolate treatment effects while controlling for country-level fixed effects.

**Main study: the prolific experiment**

**Procedures.** The Main Study experiment was programmed using oTree (Chen et al. 2016) and conducted online via the Prolific platform between August and December 2024, drawing participants from Italy, Denmark, and the UK. While maintaining consistency with the Pilot Study’s core experimental framework and methodology, in the Main Study we introduced four key

**Table 3 Participants' attitudes towards immigrants (pilot study).**

<i>Immigrants take jobs away from natives in a country</i>	Freq.	%	Cum.%
Disagree strongly	137	50.183	50.183
Disagree	93	34.066	84.249
Neutral	34	12.454	96.703
Agree	6	2.198	98.901
Agree strongly	3	1.099	100
<i>Immigrants make crime problems worse</i>	Freq.	%	Cum.%
Disagree strongly	55	20.147	20.147
Disagree	88	32.234	52.381
Neutral	58	21.245	73.626
Agree	57	20.879	94.505
Agree strongly	15	5.495	100
<i>Immigrants are a strain on the [country] welfare system</i>	Freq.	%	Cum.%
Disagree strongly	56	20.513	20.513
Disagree	93	34.066	54.579
Neutral	62	22.711	77.289
Agree	45	16.484	93.773
Agree strongly	17	6.227	100
<i>Today in [country], there are too many immigrants</i>	Freq.	%	Cum.%
Disagree strongly	65	23.810	23.810
Disagree	87	31.868	55.678
Neutral	72	26.374	82.051
Agree	35	12.821	94.872
Agree strongly	14	5.128	100
Total	273	100	

**Table 4 Summary statistics of the experimental variables (Pilot Study).**

Variable	Mean	SD	Median	Min	Max
Correct counting “a”s in the paragraph (0/1)	0.143	0.351	0	0	1
Points earned in the real-effort task	83.755	17.520	88	1	100
Points reported	60.575	35.610	73	0	100
Compliance rate	0.720	0.391	1	0	1
“Taxes” paid by the player before being audited	18.173	10.683	21.900	0	30
1 if the player was audited; 0 otherwise	0.040	0.197	0	0	1
If audited, 1 for true reporting; 0 for fraud	0.727	0.467	1	0	1
Final earnings in points (after “taxes” and audits)	65.151	17.094	65.800	1	100
Total	273				

Q6

**Table 5 OLS regressions of treatment effects on compliance rate (pilot study).**

DV: compliance rate	(1)	(2)	(3)	(4)	(5)
Natives Treat. (Nat.)	-0.002 (0.059)	0.006 (0.058)	0.015 (0.095)	0.005 (0.066)	-0.008 (0.067)
Immigrants Treat. (Imm.)	0.027 (0.060)	0.039 (0.057)	-0.010 (0.096)	0.051 (0.069)	0.044 (0.067)
Italy (IT)		0.227*** (0.045)	0.200* (0.089)		0.166* (0.083)
Nat. × IT			-0.019 (0.116)		
Imm. × IT			0.100 (0.115)		
Constant	0.712*** (0.045)	0.593*** (0.053)	0.607*** (0.073)	1.199*** (0.313)	1.231*** (0.333)
Controls				X	X
Additional Controls					X
H <sub>0</sub> <sup>a</sup>					
$\beta_{Nat} = \beta_{Imm}$	0.610	0.543	0.777	0.455	0.394
$\beta_{Nat \times IT} = \beta_{Imm \times IT}$			0.255		
Observations	273	273	273	241	241
R <sup>2</sup>	0.001	0.086	0.090	0.112	0.172
AIC	266.250	244.154	246.764	252.101	253.362
BIC	277.079	258.592	268.421	318.312	350.937

Notes: OLS regressions, with robust standard errors in parentheses. In all models, the dependent variable is compliance rate, defined as the ratio between reported to earned points. Baseline: Control treatment (omitted). Models 4 and 5 include the following set of controls: gender (1 for males); age; dummies for each education level; dummies for each type of employment status; having Economics as an educational background; the approximate monthly household income, after “taxes” and other deductions; having participated in other experimental studies; the amount of failed comprehension questions; the amount of “a”s in the paragraph. Model 5 adds the following set of additional controls: political stance (1 for right; 0 otherwise); belonging to and/or having contacts with any humanitarian or charitable organization; frequency of attending religious services, apart from weddings, funerals, and christenings; risk preferences (1 for avoiding risks; 0 for taking risks); attitudes towards immigrants. For brevity reasons, we have used the “X” symbol to indicate that the regression model includes control variables. We refer to Supplementary Appendix A Table S2 online for the full specification of each control variable.

<sup>a</sup>Wald tests hypotheses on the parameters of the regression model; p-values reported in italics.

\*p < 0.10, \*\*p < 0.05, \*\*\*p < 0.01, \*\*\*\*p < 0.001.

**Table 6 Number of observations, by treatment and country (Main Study).**

Treatment	Italy	Denmark	UK	Pooled sample
Control Treat.	306	22	22	350
	32.69%	36.67%	34.38%	33.02%
Natives Treat.	320	17	20	357
	34.19%	28.33%	31.25%	33.68%
Immigrants Treat.	310	21	22	353
	33.19%	35.00%	34.38%	33.30%
Total	936	60	64	1060

modifications aimed at enhancing sample representativeness and statistical power.

First, to enhance sample representativeness, in the Main Study we implemented stricter eligibility criteria, ensuring that participants were not only natives, but also had direct experience with paying taxes on their earned incomes. Specifically, participants had to meet all three of the following eligibility requirements, with the first two being identical to those in the Pilot Study (ensuring a native sample), while the third was newly introduced in the Main Study:

- Being born in the country of participation (Italy/Denmark/the UK);
- Having both parents born in the country of participation;
- Paying income tax—or having paid them in the past—in the country of participation.

Participants who failed to meet any of the three requirements were screened out but still compensated for their time, as required by the Prolific pre-screening rules. This recruitment approach ensured that the final sample was more representative of taxpayers, reducing reliance on student participants. Indeed, while the Prolific sample still includes a mix of employment backgrounds, it now consists primarily of working adults; importantly, all participants have paid or are currently paying

income taxes, making the sample more representative and our findings more relevant for policy implications.

Second, to increase overall statistical power, we recruited the maximum number of active and eligible participants in Italy and Denmark available on Prolific, within the limits of our research budget. Compared to the Pilot Study (with 135 Italians and 138 Danes), in the Main Study we successfully recruited 936 Italians—an increase of nearly seven-fold. But a total of only 260 Danish users were registered on Prolific, and after applying our strict eligibility filters, we could recruit just 60 participants. In an effort to augment the sample size and improve the generalizability of our findings, we included a small sample of individuals from the UK (64 participants), as they were the sample best aligned with our study’s aims and most readily available on Prolific.

Third, the experimental instructions and interface in the Main Study were fully translated into the native languages of the participants (Italian for Italians; Danish for Danes; English for UK participants). This contrasts with the Pilot Study, in which the instructions were provided in English for Danish participants.<sup>15</sup> This modification minimized misunderstandings and enhanced clarity, ensuring that all participants fully understood the instructions, while also reducing possible response biases linked to linguistic ambiguity.

Fourth, to address potential concerns about the authenticity of the donations, in the Main Study we incorporated additional transparency measures. For example, participants were explicitly informed in the instructions that we could provide them an official certificate from the charity—as well as bank transfer receipts—confirming the allocation of funds to the specific, targeted recipients (unemployed people/natives/immigrants). No participants requested such documentation, suggesting confidence in the legitimacy of the donation process.

Table 6 presents the number of observations per treatment and country in the Main Study, with a total of 1060 participants (936 from Italy, 60 from Denmark, and 64 from the UK). Compared to the Pilot Study, the distribution of participants across treatments is more evenly balanced—approximately one-third assigned to each treatment group, as shown in Table 6 (vs. Table 1).

**Table 7 Participants' demographics and other characteristics (main study).**

Variable	Mean	SD	Median	Min	Max
Italy	0.883	0.322	1	0	1
Denmark	0.057	0.231	0	0	1
UK	0.060	0.238	0	0	1
Sex (1 male)	0.571	0.495	1	0	1
Age	33.465	10.426	30	18	77
Student (0/1)	0.197	0.398	0	0	1
Economist (0/1)	0.096	0.295	0	0	1
Have kids (0/1)	0.179	0.384	0	0	1
Participated in other similar studies (0/1)	0.358	0.480	0	0	1
- If yes, how many	3.652	2.296	3	1	7
Amount of failed comprehension questions	0.708	0.899	0	0	6
<i>Education level</i>	Freq.	%	Cum.%		
Junior High School	15	1.415	1.415		
High School	360	33.962	35.377		
Bachelor Degree	266	25.094	60.472		
Master Degree	266	25.094	85.566		
Post Graduate	31	2.925	88.491		
Specialization					
Master	68	6.415	94.906		
Ph.D.	47	4.434	99.340		
Other	7	0.660	100		
<i>Monthly household income (in EUR)</i>	Freq.	%	Cum.%		
<EUR 3k	525	52.977	52.977		
EUR 3k to 6k	351	35.419	88.396		
EUR 6k to 10k	77	7.770	96.165		
>EUR 10k	38	3.835	100		
Total	991	100			
<i>Political stance</i>	Freq.	%	Cum.%		
Left or Center	843	79.528	79.528		
Right	217	20.472	100		
<i>Belong to and/or have contacts with any humanitarian or charitable organization</i>	Freq.	%	Cum.%		
No	796	75.094	75.094		
Yes	264	24.906	100		
<i>Frequency of attending religious services (apart from weddings, funerals, and christenings)</i>	Freq.	%	Cum.%		
More than once a week	12	1.132	1.132		
Once a week	40	3.774	4.906		
Once a month	29	2.736	7.642		
Only on specific holy days	169	15.943	23.585		
Once a year	60	5.660	29.245		
Practically never	750	70.755	100		
<i>How do you see yourself: Are you in general a person who takes risks or do you try to avoid risks?</i>	Freq.	%	Cum.%		
Take risk	292	27.547	27.547		
Avoid risk	768	72.453	100		
Total	1060	100			

**Sample.** Table 7 reports the statistics for the pooled sample (N = 1060); see Supplementary Appendix A Table S1 online for additional descriptives.

Overall, the majority of our participants (69.15%) were employed and currently paying income taxes. Another 11.13% were retired, unemployed, or other status, but all were paying income taxes or have paid them in the past. Only 19.72% of the sample were students, compared to 52.75% in the Pilot Study. The sample was 57.10% male, with an average age of 33.47 years (ranging from 18 to 77 years). Regarding education, the largest group of participants held a High School Diploma (33.96%), followed by those with a Bachelor's Degree (25.09%) and a Master's Degree (25.09%). Monthly household income data revealed that 52.98% earned less than EUR 3000, while only

**Table 8 Participants' attitudes towards immigrants (main study).**

Immigrants take jobs away from natives in a country	Freq.	%	Cum.%
Disagree strongly	459	43.302%	43.302%
Disagree	355	33.491%	76.792%
Neutral	172	16.226%	93.019%
Agree	55	5.189%	98.208%
Agree strongly	19	1.792%	100%
<i>Immigrants make crime problems worse</i>	Freq.	%	Cum.%
Disagree strongly	148	13.962%	13.962%
Disagree	272	25.660%	39.623%
Neutral	273	25.755%	65.377%
Agree	272	25.660%	91.038%
Agree strongly	95	8.962%	100%
<i>Immigrants are a strain on the [country] welfare system</i>	Freq.	%	Cum.%
Disagree strongly	199	18.774%	18.774%
Disagree	292	27.547%	46.321%
Neutral	287	27.075%	73.396%
Agree	207	19.528%	92.925%
Agree strongly	75	7.075%	100%
<i>Today in [country], there are too many immigrants</i>	Freq.	%	Cum.%
Disagree strongly	227	21.415%	21.415%
Disagree	291	27.453%	48.868%
Neutral	292	27.547%	76.415%
Agree	164	15.472%	91.887%
Agree strongly	86	8.113%	100%
Total	1060	100	

3.83% reported earnings exceeding EUR 10,000. In terms of political orientation, 20.47% leaned right; additionally, 24.91% reported some affiliation with humanitarian or charitable organizations; and 7.64% said they attended religious services (other than weddings, funerals and christenings) at least once a month. Regarding risk preferences, 27.55% described themselves as risk-takers, whereas 72.45% preferred to avoid risks. Beyond these core demographics, 35.8% of participants had previously participated in similar studies, with a median of three prior participations. Additionally, the mean number of failed comprehension questions was 0.71, indicating a high level of understanding of the experimental tasks.

Table 8 presents summary statistics on participants' attitudes toward immigrants. As in the Pilot Study (Table 3), participants generally reported low levels of anti-immigrant sentiment, particularly regarding economic concerns.<sup>16</sup> A majority (76.79%) disagreed or strongly disagreed with the statement that immigrants take jobs away from natives (the figure was 84.25% in the Pilot Study). Similarly, almost half of participants (46.32%) disagreed that immigrants place a strain on the welfare system (54.58% in the Pilot Study). Attitudes were more divided on issues related to crime and overall immigration levels. On whether immigrants make crime problems worse, responses were nearly evenly split, with 39.62% disagreeing, 25.76% neutral, and 34.62% agreeing. Opinions on whether there are "too many immigrants" in the country were more polarized, with 48.87% disagreeing, 27.55% neutral, and 23.58% agreeing.

**Results.** Table 9 presents descriptive statistics on key experimental variables measured during the real-effort task and reporting stage.

Regarding the real-effort task, 185 out of 1060 participants (18.5%) correctly counted the "a"s in the paragraph, earning the maximum possible points. On average, participants earned 84.73

**Table 9 Summary statistics of the experimental variables (Main Study).**

Variable	Mean	SD	Median	Min	Max
Correct counting “a”s in the paragraph (0/1)	0.185	0.388	0	0	1
Points earned in the real-effort task	84.733	17.698	90	1	100
Points reported	70.731	30.467	82	0	100
Compliance rate	0.837	0.304	1	0	1
“Taxes” paid by the player before being audited	21.219	9.140	25	0	30
1 if the player was audited; 0 otherwise	0.049	0.216	0	0	1
If audited, 1 for true reporting; 0 for fraud	0.673	0.474	1	0	1
Final earnings in points (after “taxes” and audits)	63.123	15.846	65	0.700	100
Total	1,060				

points, with a median of 90 points, and reported an average of 70.73 points, resulting in an average compliance rate of 0.84 (versus 0.72 in the Pilot Study).<sup>17</sup> Fifty-two participants (4.9%) were randomly audited. Among them, 67.3% truthfully reported their earnings, while the remaining 32.7% underreported and were fined.

Table 10 presents OLS regression estimates assessing treatment effects on compliance rates, incorporating country-fixed effects. As in the Pilot Study (Table 5), results indicate no statistically significant differences in compliance across treatments, confirming the robustness of null effects.

Table 10 is organized to mirror as closely as possible the structure of Table 5 from the Pilot Study, with additional regression models denoted by the suffixes “b” and “c.” All specifications include treatment dummies, with the Control group as the reference category. Consistent with Table 5, Models (2), (3), and (5) introduce the Italy dummy, and Model (3) additionally includes interaction terms between the Italy dummy and each treatment dummy. As extensions, Models (2b), (3b), (5b), and (5c) incorporate fixed effects for all three countries—Italy, Denmark, and the UK—while Model (3b) further includes interactions between the treatment dummies and the full set of country dummies. As in the Pilot Study, demographic controls are added in all Models (4) and (5), and attitudinal controls are introduced in Models (4b), (5), and (5c).

Regression results confirm that neither the Natives nor the Immigrants treatment significantly affects compliance rates. Across all models, the treatment coefficients remain small and not statistically significant (all  $p > 0.10$ ). The inclusion of country-fixed effects does not alter this conclusion, and Wald tests confirm no significant differences between the treatment groups (all  $p > 0.10$ ).<sup>18</sup> The interaction terms between treatment groups and country dummies (Model 3b) further confirm that treatment effects do not vary significantly across countries. None of the interaction terms are statistically significant, and the Wald tests fail to reject the null hypothesis. These findings are robust to alternative model specifications, including models incorporating additional demographic and attitudinal controls (Models 4 to 5c), and subsample analyses by country and gender (available upon request).

To assess whether the lack of significant treatment effects represents true equivalence rather than a lack of statistical power, we employed the Two One-Sided Tests (TOST) procedure for unpaired data with unequal variances (Lakens, 2017; Dinno, 2024), using the Welch’s t-tests option to compare treatment pairs (Control vs. Natives, Control vs. Immigrants, Natives vs. Immigrants). We report results for an equivalence bound of  $\pm 0.05$ , corresponding to a five percentage point difference in the compliance rate between treatments. This margin is empirically grounded in the observed effect sizes from our Pilot Study (Cohen’s  $d$  ranging from  $-0.136$  to  $+0.047$ ), and is in line with, or more conservative than the bounds used in prior contributions

(see, e.g., Lakens, 2017; Grodeck and Schoenegger, 2023).<sup>19</sup> The TOST results confirm equivalence across all specifications in the pooled sample, rejecting the null hypothesis of a difference exceeding the bound between treatment groups.<sup>20</sup> These findings suggest that treatment effects are either null or smaller than a five percentage point change in compliance, and therefore practically negligible.<sup>21</sup>

**Discussion and conclusion**

According to our experimental studies, individuals’ support for redistribution—captured through simulated “tax” compliance decisions—is not statistically affected by the immigration status of the beneficiaries. This finding holds consistent across our two experiments, two distinct time periods, and three national contexts (Italy, Denmark, and the UK). Moreover, it is robust to alternative model specifications and the inclusion of a rich set of control variables. Our null result runs counter to the expectations established by previous correlational studies on welfare chauvinism and support for redistribution. With respect to theory, it raises questions about the reasons that experimental evidence diverges from findings in the social psychology literature on “in-group” favoritism and preferences for redistribution (e.g., Reeskens and Van der Meer, 2019). We hope that our work will motivate further inquiry into these discrepancies.

It is noteworthy that our findings remained consistent across the Pilot and Main studies despite three major world-historical disruptions that theory would suggest could alter attitudes toward both welfare state redistribution and immigration. Between the Pilot Study in 2020, and the Main Study in 2024, the participants experienced or observed: (1) most of the COVID-19 pandemic, which altered many people’s support for redistribution and welfare policies (Cappelen et al. 2021); (2) Russia’s invasion of Ukraine in February 2022, which created the largest forced migration in Europe since World War II; and (3) the increasing power of anti-immigration policies and coalitions in Italy, Denmark in the UK.<sup>22</sup> That the outcomes of our experiments remained stable following those disruptions offers reason for confidence in the validity of our findings.

That said, while our experimental approach improves upon prior correlational studies by introducing random treatments assignment, a controlled setting, and some features aimed at simulating real-world tax compliance decisions (e.g., real-effort task, real payments, redistribution through real donations via charity), it nonetheless remains a simulation. Therefore, as is standard with any experimental research, some caution is warranted when extrapolating findings to real-world behavior. As such, we call for both replication efforts and further evidence to be gathered under different conditions.

For example, in our experiments, most people were fully compliant with their obligation to report income earned from the clerical task—which is desirable from a social perspective, but quite “unfortunate” from a scientific perspective, since it gave us

**Table 10 OLS regressions of treatment effects on compliance rate (Main Study).**

DV: compliance rate	(1)	(2)	(2b)	(3)	(3b)	(4)	(4b)	(5)	(5b)	(5c)
Natives Treat. (Nat.)	0.004 (0.023)	0.002 (0.023)	0.002 (0.023)	0.026 (0.076)	-0.001 (0.024)	-0.023 (0.042)	-0.029 (0.039)	-0.035 (0.038)	-0.029 (0.041)	-0.035 (0.038)
Immigrants Treat. (Imm.)	-0.004 (0.023)	-0.005 (0.023)	-0.005 (0.023)	0.017 (0.073)	-0.008 (0.024)	-0.056 (0.040)	-0.047 (0.038)	-0.048 (0.037)	-0.058 (0.039)	-0.049 (0.037)
Italy (IT)		0.101** (0.032)		0.118* (0.052)				0.229*** (0.057)		
Denmark (DK)			-0.117** (0.044)		-0.140+ (0.081)				-0.269*** (0.078)	-0.266*** (0.070)
UK			-0.087* (0.044)		-0.096 (0.062)				-0.178* (0.088)	-0.177* (0.088)
Nat. x IT				-0.027 (0.079)						
Imm. x IT				-0.025 (0.076)						
Nat. x DK					0.047 (0.117)					
Nat. x UK					0.006 (0.102)					
Imm. x DK					0.028 (0.104)					
Imm. x UK					0.021 (0.106)					
Constant	0.837*** (0.016)	0.748*** (0.033)	0.850*** (0.016)	0.734*** (0.050)	0.852*** (0.017)	0.533* (0.259)	0.664** (0.246)	0.384 (0.250)	0.439+ (0.255)	0.605* (0.238)
Controls						X	X	X	X	X
Additional Controls			X		X					
Country Fixed Effect										
H <sub>0</sub> <sup>a</sup>										
$\beta_{Nat} = \beta_{Imm}$	0.699	0.759	0.762	0.9089	0.7702	0.3899	0.6296	0.7166	0.4400	0.7145
$\beta_{Nat*IT} = \beta_{Imm*IT}$				0.9809						
$\beta_{Nat*DK} = \beta_{Imm*DK}$					0.8607					
$\beta_{Nat*UK} = \beta_{Imm*UK}$					0.9037					
Observations	1060	1060	1060	1060	1060	356	356	356	356	356
R <sup>2</sup>	0.000	0.012	0.012	0.012	0.012	0.136	0.199	0.241	0.182	0.243
AIC	490.217	479.977	481.686	483.796	489.416	176.628	165.746	148.428	160.924	149.596
BIC	505.115	499.841	506.516	513.592	534.110	265.751	285.869	272.426	257.798	277.469

Notes: OLS regressions, with robust standard errors in parentheses. The dependent variable is compliance rate, defined as the ratio of reported to earned points. The Control treatment serves as the baseline category, with Natives and Immigrants treatments included as treatment dummies. Model (2), (3), and (5) include the Italy dummy; Models (2b), (3b), (5b), and (5c) incorporate all country-fixed effects (Denmark, UK, and Italy as baseline hence omitted). Models (3) and (3b) introduce interaction terms between treatment and country dummies. Models (4) to (5c) add demographic controls, such as gender, age, education level, employment status, household income, and prior experimental participation; Models (4b), (5) and (5c) further includes attitudinal controls (political orientation, religious affiliation, humanitarian affiliations, risk preferences). For brevity reasons, we have used the "X" symbol to indicate that the regression model includes control variables and/or country fixed effect. We refer to Supplementary Appendix A Table S2 online for the full specification of each variable.

<sup>a</sup>Wald tests hypotheses on the parameters of the regression model; *p* values reported in italics.

\**p* < 0.10, \*\**p* < 0.05, \*\*\**p* < 0.001.

too little variation in compliance. To address this issue, future extensions could vary the cost-benefit of income reporting requirements. Indeed, an alternative hypothesis to our research design is that people may break their “moral code” and evade taxes due to anti-immigration attitudes only when the personal benefits are sufficiently greater than the related costs.

Additionally, our research design—using the standard tax evasion game (Alm and Malézieux, 2020) established and applied over decades of research—has some potential drawbacks that deserve further investigation. For example, given that participants’ final earnings depend on both their effort in the clerical task and their honesty in reporting income from that task, there may be a hedging problem: if participants realize the real-effort task is too difficult, they could decide to exert less effort, since they later can under-report their task performance to recoup some earnings, even if their true preferences were not to evade. Relatedly, informing participants about the two stages (real-effort task and reporting stage) may create confusion as to whether the treatment results are generated by the reporting or the effort channel, or by a combination of the two. For example, it might be that knowing ex-ante that the benefits will go to immigrants changes participants’ effort levels rather than their income reporting decisions. To address these issues, the process of reporting earnings for “tax” purposes could be introduced after the real-effort task, but without informing the participants in advance. Similarly, to address potential demand effects, participants could be informed of “tax” and audit rules they will face only *after* they earn income. Although this would be inconsistent with real-world tax systems, and might introduce unexpected losses for participants (which some research suggests could influence attitudes toward redistribution, e.g., Cappelen et al. 2021), future studies could explore variations in disclosure timing and how those affect compliance and fairness perceptions.

To conclude, our two studies advance social scientific knowledge by examining a classic social dilemma, in which individuals find themselves at the center of a conflict between self and society, as well as between natives and immigrants. It links two literatures that have developed independently and are usually not in conversation with one another: economic research on redistribution, and the stream of research on welfare chauvinism, conducted mostly in political science. Our findings, as well as the limitations inherent in experimental approaches, open several promising avenues for future research on immigration, redistribution, and the welfare state.

### Data availability

Data available upon request.

Received: 30 April 2024; Accepted: 14 July 2025;

Published online: xx xxx 2025

### Notes

- 1 Among others, Tabellini (2020) shows that immigrants are perceived as a fiscal burden; this finding is consistent with a large literature showing that ethnic diversity is associated with lower public goods provision and lower support for redistribution (e.g., Alesina et al. 1999; Luttmer, 2001).
- 2 For similar approaches in the tax evasion experimental literature, see also Alm et al. (1993), Christian and Alm (2014), Zhang et al. (2016), and Pampel et al. (2019).
- 3 The experimental instructions are available in Supplementary Appendix C online.
- 4 In the Pilot Study, we applied country-specific point conversion rates (1 point = 0.13 EUR in Italy; 1.34 DKK in Denmark) to approximate local purchasing power. In the Main Study, we adopted a uniform conversion rate (1 point = US\$0.02) across all countries to align with Prolific’s payment structure, which encourages uniformity in compensation based solely on study duration, regardless of participant location or nationality. We used US\$ because Prolific currently supports payments only in

British Pounds and US Dollars—and our research funds were received in US Dollars and credited to our Prolific account in that currency.

- 5 To test comprehension, at the end of the instructions and before starting the real-effort task, participants were asked a series of questions about the game. They were provided with prompt feedback via computer. In case they gave an incorrect answer, the feedback provided an explanation of the mistake and the correct answer. In the regression analyses, we control for an individual’s number of failed questions, which also includes the unlikely but possible repetitions of an error.
- 6 The number of points earned in the real-effort task are computed as  $100 \times \max(0, 1 - |R - T|/T)$ , where  $T$  is the true number of “a”s in the paragraph, and  $R$  is the participant’s reported count. Participants receive the maximum score (100 points) if they report the exact correct number ( $R = T$ ). If the reported count deviates from the correct value, the score decreases proportionally. If the reported number is zero, or more than double the correct count ( $R \geq 2T$ ), the participant receives zero points. For example, if  $T = 50$  and the participant reports  $R = 49$ , the score is 98 points, whereas if  $R = 1$ , the score is 2 points.
- 7 If time expired and the participant had not provided any answer, s/he got zero points and was automatically logged out from the study without any payment. This was necessary because otherwise s/he would have had no points to report in the next task.
- 8 In the instructions, we generally referred to donation to a charity, while deliberately avoiding any explicit mention of the Red Cross. This is because individuals may have different relationships with and attitudes towards particular charitable organizations, which may affect their behavior.
- 9 In the Pilot Study, subjects were paid via PayPal, with an average payment of EUR 8.23 in Italy and DKK 91.14 in Denmark, including a minimum payment for completing the study of EUR 5 in Italy and DKK 37.06 in Denmark. In the Main Study, subjects were paid via Prolific, with an average payment of US\$3.27 in Italy, US\$3.30 in Denmark, and US\$3.18 in the UK; US\$2 was the minimum payment for completing the study.
- 10 Our results in the studies reported here are consistent with those previous findings (Guerra and Harrington, 2018; Pampel et al. 2019), in that “tax” compliance rates (within experiments) are higher in Italy than in Denmark and the UK.
- 11 Political stance was originally measured on a 1–10 scale, where 1 = Left and 10 = Right (see Supplementary Appendix A Table S1 online). For analytical purposes, we categorized participants as Right (coded as 1) if they selected 7 or higher, and as 0 otherwise. This dichotomization follows existing literature linking right-leaning political views to attitudes toward immigration (Mudde, 2019; Shehaj et al. 2019). As a preview of our results (on which there are more details in the next sections), 40.85% of participants in the Main Study described themselves as Left-leaning, while 38.68% identified as Centrists. In the Pilot Study, political orientation followed a similar pattern, with a majority of participants identifying as Left or Center. This ideological skew is not surprising. As is common in online experimental research, the sample tends to overrepresent younger and well-educated individuals, who are more likely to have progressive political views (e.g., Berinsky et al. 2012; Coppock and McClellan, 2019). While this may not fully reflect the political composition of the general population, the consistency of our findings across different time periods and distinct samples—collected on different platforms with different demographic profiles—strengthens confidence in the robustness of our results.
- 12 A series of balance tests (Supplementary Appendix B online)—conducted for both studies, on both the pooled sample and each country-specific sample, with p-values computed following Chiappello (2018)—confirm that the samples are balanced across treatments in terms of personal characteristics and other relevant preferences. A few tests became statistically significant due to the relatively low number of observations in specific categories in some treatments.
- 13 Each announcement included the following general text: “Our research lab is looking for 300 volunteers in Italy/Denmark to take a short and paid online survey as part of a research project designed by researchers from the Copenhagen Business School. It is short, safe, and anonymous, and all participants will get a minimum payment for their contribution. Click here to register.”
- 14 The Italy dummy is positive and statistically significant, indicating that—consistent with previous findings (Guerra and Harrington, 2018)—tax compliance rates are higher in Italy than in Denmark ( $\beta = 0.227$ ,  $p < 0.001$  in Model 2).
- 15 Regarding the potential that English-language instructions for Danish participants might have introduced bias into our Pilot Study results, we note that the University of Copenhagen estimates that 86% of Danes are fluent in English, due to mandatory English language instruction from early elementary school onward; in addition, official Danish government documents are routinely published in English and Danish simultaneously. English is so well incorporated into everyday Danish life, regardless of social class or educational level, that there is little reason for concern about the English language instructions used in the Pilot Study.
- 16 Overall, both studies indicate low levels of perceived economic and welfare-related threats from immigration. The Main Study reveals greater neutrality and a slight shift toward more concern about crime and immigration levels compared to the Pilot Study. This slight difference may stem from the larger and more diverse sample in the Main Study, which included a broader range of perspectives on immigration.
- 17 Overall, willingness to comply with “tax” obligations was similar across the two studies, though compliance was slightly higher in the Main Study than in the Pilot

Study. The compliance rate increased from 0.72 in the Pilot Study to 0.84 in the Main Study, and the proportion of participants declaring at least some taxable income rose from 86.1% in the Pilot Study to 95.3% in the Main Study. Additionally, the amount of “tax” paid before audits was slightly higher in the Main Study (21.22 points) compared to the Pilot Study (18.17 points).

- 18 As in the Pilot Study (Table 5), the Italy dummy in Table 10 is positive and statistically significant, indicating that—consistent with previous findings (Guerra and Harrington, 2018; Pampel et al. 2019)—experimental “tax” compliance rates are higher in Italy than in Denmark and the UK ( $\beta = 0.101, p < 0.01$  in Model 2).
- 19 Defining equivalence margins is a well-recognized challenge in applied research (see, e.g., Wiens, 2022; Hung et al. 2005). Equivalence margins vary considerably, with many studies adopting bounds between  $\pm 0.1$  and  $\pm 0.5$ . In this perspective, our choice of a  $\pm 0.05$  margin is relatively cautious, yet sufficiently powered given our sample size.
- 20 Similarly, in the Italian sample, the large number of observations provides sufficient statistical power to establish robust evidence of treatment equivalence. In contrast, the Danish and UK samples yield inconclusive results due to their small size.
- 21 The full STATA output and code are provided in the Supplementary Appendix D online.
- 22 In 2022, Italy elected a right-wing coalition led by Prime Minister Giorgia Meloni, who campaigned on a promise to create a naval blockade to stop migrant boats coming to Italy from North Africa; she has also proposed sending asylum seekers to Albania. In the same year, Denmark’s government formed around the leadership of Mette Frederiksen, a left-center Social Democrat who ran on an anti-immigrant stance, including a “zero refugee policy.” Both Denmark and the UK proposed in 2022 to send asylum seekers to Rwanda.

## References

- Alesina A, Baqir R, Easterly W (1999) Public goods and ethnic divisions. *Q J Econ* 114(4):1243–1284
- Alesina A, Glaeser E (2004) *Fighting poverty in the US and Europe: A world of difference*. Oxford University Press, Oxford, UK
- Alesina A, Miano A, Stantcheva S (2023) Immigration and redistribution. *Rev Econ Stud* 90(1):1–39
- Almås I, Cappelen AW, Tungodden B (2020) Cutthroat capitalism versus cuddly socialism: Are Americans more meritocratic and efficiency-seeking than Scandinavians? *J Political Econ* 128(5):1753–1788
- Alm J (2019) What motivates tax compliance? *J Econ Surv* 33(2):353–388
- Alm J, Malézieux A (2021) 40 years of tax evasion games: A meta-analysis. *Exp Econ* 24(3):699–750
- Alm J, Jackson BR, McKee M (1993) Fiscal exchange, collective decision institutions, and tax compliance. *J Econ Behav Organ* 22(3):285–303
- Barr A, Lane T, Nosenzo D (2018) On the social inappropriateness of discrimination. *J Public Econ* 164:153–164
- Bauböck R, Scholten P (2016) Coping with ‘the Progressive’s Dilemma’: Nationhood, Immigration and the Welfare State. *Comp Migr Stud* 4(4):1–7
- Berge LJO, Bjorvatn K, Galle S, Miguel E, Posner DN, Tungodden B, Zhang K (2020) Ethnically biased? Experimental evidence from Kenya. *J Eur Econ Assoc* 18(1):134–164
- Berinsky AJ, Huber GA, Lenz GS (2012) Evaluating online labor markets for experimental research: Amazon.com’s Mechanical Turk. *Political Anal* 20(3):351–368
- Bernhard H, Fischbacher U, Fehr E (2006) Parochial altruism in humans. *Nature* 442(7105):912–915
- Bommes M, Geddes A (2000) *Immigration and Welfare: Challenging the Borders of the Welfare State*. Routledge, London, UK
- Borisova E, Govorun A, Ivanov D, Levina I (2018) Social capital and preferences for redistribution to target groups. *Eur J Political Econ* 54:56–67
- Buss C (2019) Public opinion towards targeted labour market policies: A vignette study on the perceived deservingness of the unemployed. *J Eur Soc Policy* 29(2):228–240
- Cappelen AW, Falch R, Sørensen EØ, Tungodden B (2021) Solidarity and fairness in times of crisis. *J Econ Behav Organ* 186:1–11
- Christian RC, Alm J (2014) Empathy, sympathy, and tax compliance. *J Econ Psychol* 40:62–82
- Coppock A, McClellan OA (2019) Validating the demographic, political, psychological, and experimental results obtained from a new source of online survey respondents. *Res Politics* 6(1):1–14
- Cullen JB, Turner N, Washington E (2021) Political alignment, attitudes toward government, and tax evasion. *Am Econ J: Econ Policy* 13(3):135–166
- Currarini S, Mengel F (2016) Identity, homophily and in-group bias. *Eur Econ Rev* 90:40–55
- Dahlberg M, Edmark K, Lundqvist H (2012) Ethnic diversity and preferences for redistribution. *J Polit Econ* 120(1):41–76
- Dinno, A 2024. *tostt: Mean-equivalence t tests*. In: *tost Stata software package*. Retrieved from: <https://www.alexisdinno.com/stata/tost.html>
- Doerrenberg P, Duncan D (2014) Experimental evidence on the relationship between tax evasion opportunities and labor supply. *Eur Econ Rev* 68:48–70
- Doerrenberg P (2015) Does the use of tax revenue matter for tax compliance behavior? *Econ Lett* 128:30–34
- Eger M (2010) Even in Sweden: The Effect of Immigration on Support for Welfare State Spending. *Eur Sociol Rev* 26(2):203–217
- Enke B, Rodriguez-Padilla R, Zimmermann F (2022) Moral universalism: Measurement and economic relevance. *Manag Sci* 68(5):3590–3603
- Felfe C, Kocher MG, Rainer H, Saurer J, Siedler T (2021) More opportunity, more cooperation? The behavioral effects of birthright citizenship on immigrant youth. *J Public Econ* 200:104448
- Fershtman C, Gneezy U (2001) Discrimination in a segmented society: An experimental approach. *Q J Econ* 116(1):351–377
- Finseraas H (2008) Immigration and preferences for redistribution: An empirical analysis of European survey data. *Comp Eur Politics* 6(4):407–431
- Fong CM, Luttmer EF (2011) Do fairness and race matter in generosity? Evidence from a nationally representative charity experiment. *J Public Econ* 95(5):372–394
- Fouka V, Tabellini M (2022) Changing in-group boundaries: the effect of immigration on race relations in the United States. *Am Political Sci Rev* 116(3):968–984
- Gilens M (2000) *Why Americans Hate Welfare: Race, Media, and the Politics of Antipoverty Policy*. The University of Chicago Press, Chicago, IL
- Goette L, Huffman D, Meier S (2006) The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups. *Am Econ Rev* 96(2):212–216
- Grodeck B, Schoenegger P (2023) Demanding the morally demanding: Experimental evidence on the effects of moral arguments and moral demandingness on charitable giving. *J Behav Exp Econ* 103:101988
- Guerra A, Harrington B (2018) Attitude–behavior consistency in tax compliance: A cross-national comparison. *J Econ Behav Organ* 156:184–205
- Guerra, A, & Harrington, B (2021). *Why do People Pay Taxes? Explaining Tax Compliance by Individuals*, in *Handbook on The Politics of Taxation*, Lukas Hakeberg and Laura Seelkopf (Eds.), Chapter 23, pages 356–374. Cheltenham, UK: Edward Elgar Publishing. ISBN: 9781788979412
- Guerra A, Harrington B (2023) Regional variation in tax compliance and the role of culture. *Economia Politica* 40(1):139–152
- Huff C, Tingley D (2015) “Who are these people?” Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Res Politics* 2(3):1–12
- Hung H, Wang S-J, O’Neill R (2005) A regulatory perspective on choice of margin and statistical inference issue in non-inferiority trials. *Biometrical J* 47(1):28–36
- Jørgensen MB, Thomsen TL (2016) Deservingness in the Danish context: Welfare chauvinism in times of crisis. *Crit Soc Policy* 36(3):330–351
- Kim D, Lee D (2021) Immigration and the pattern of public spending: Evidence from OECD countries. *Int Tax Public Financ* 28(4):1014–1034
- Kootstra A (2017) Deserving and Undeserving Welfare Claimants in Britain and the Netherlands: Examining the Role of Ethnicity and Migration Status Using a Vignette Experiment. *Eur Sociol Rev* 32(3):325–339
- Lago-Peñas I, Lago-Peñas S (2010) The Determinants of Tax Morale in Comparative Perspective: Evidence from European Countries. *Eur J Political Econ* 26(4):441–453
- Lakens D (2017) Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Soc Psychol Personal Sci* 8(4):355–362
- Lane T (2016) Discrimination in the laboratory: A meta-analysis of economics experiments. *Eur Econ Rev* 90:375–402
- Lewis A, Carrera S, Cullis J, Jones P (2009) Individual, cognitive and cultural differences in tax compliance: UK and Italy compared. *J Econ Psychol* 30(3):431–445
- Luttmer EF (2001) Group loyalty and the taste for redistribution. *J Polit Econ* 109(3):500–528
- Mudde C (2019) *The Far Right Today*. Wiley, New York
- Nemore F, Morone A (2019) Public spirit on immigration issues and tax morale in Italy: An empirical investigation. *J Behav Exp Econ* 81:11–18
- Nicholls WJ (2020) The dark side of immigrant integration policies: Day laborers in the United States. *Soc Probl* 67(3):437–451
- Ottone S, Ponzano F, Andrighetto G (2018) Tax compliance under different institutional settings in Italy and Sweden: an experimental analysis. *Economia Politica* 35:367–402
- Pampel F, Andrighetto G, Steinmo S (2019) How institutions and attitudes shape tax compliance: a cross-national experiment and survey. *Soc Forces* 97(3):1337–1364
- Paolacci G, Chandler J, Ipeirotis PG (2010) Running experiments on Amazon Mechanical Turk. *Judgm Decis Mak* 5(5):411–419
- Reeskens, T & van der Meer, T (2017). *The Relative Importance of Welfare Deservingness Criteria*, in W van Oorschot, F Roosma, B Meuleman and T Reeskens (eds). *The Social Legitimacy of Targeted Welfare: Attitudes to Welfare Deservingness*, pp. 55–72. Cheltenham, UK: Edward Elgar Publishing

- Reeskens T, Van der Meer T (2019) The inevitable deservingness gap: A study into the insurmountable immigrant penalty in perceived welfare deservingness. *J Eur Soc Policy* 29(2):166–181
- Reeskens T, Van Oorschot W (2012) Disentangling the ‘New Liberal Dilemma’: On the relation between general welfare redistribution preferences and welfare chauvinism. *Int J Comp Sociol* 53(2):120–139
- Rodriguez-Justicia D, Theilen B (2022) Immigration and tax morale: the role of perceptions and prejudices. *Empir Econ* 62(4):1801–1832
- Shehaj A, Shin A, Inglehart R (2019) Immigration and right-wing populism: An origin story. *Part Politics* 27(2):282–293
- Tabellini M (2020) Gifts of the immigrants, woes of the natives: Lessons from the age of mass migration. *Rev Econ Stud* 87(1):454–486
- Van der Meer T, Reeskens T (2021) Welfare chauvinism in the face of ethnic diversity: A vignette experiment across diverse and homogenous neighborhoods on the perceived deservingness of native and foreign-born welfare claimants. *Eur Sociol Rev* 37(1):89–103
- Van der Waal J, Achterberg P, Houtman D, De Koster W, Manevska K (2010) ‘Some are more equal than others’: Economic egalitarianism and welfare chauvinism in the Netherlands. *J Eur Soc Policy* 20(4):350–363
- Van Der Waal J, De Koster W, Van Oorschot W (2013) Three worlds of welfare chauvinism? How welfare regimes affect support for distributing welfare to immigrants in Europe. *J Comp Policy Anal: Res Pract* 15(2):164–181
- van Oorschot W (2006) Making the difference in social Europe: Deservingness perceptions among citizens of European welfare states. *J Eur Soc Policy* 16(1):23–42
- Wiens BL (2002) Choosing an equivalence limit for noninferiority or equivalence studies. *Controlled Clin Trials* 23(1):2–14
- Zhang N, Andrighetto G, Ottone S, Ponzano F, Steinmo S (2016) “Willing to pay?” Tax compliance in Britain and Italy: An experimental analysis. *PLoS one* 11(2):e0150277

## Acknowledgements

The Pilot Study received funding from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 748062 “ChEATAX,” for which Alice Guerra is the grant holder. The Main Study was supported by the authors’ research funds from the Department of Economics at the University of Bologna (Alice Guerra) and the Rockefeller Center for Public Policy at Dartmouth College (Brooke Harrington). We are grateful to Romain Espinosa, Etienne Lehmann, Peter Marsden, Nadine Riedel, Gerd Muehlheusser, and Rupert Sausgruber for their extremely valuable feedback. For helpful comments, we also thank Pedro M. Barros, Emanuela Carbonara, Marco Casari, John D’Attoma, Oliver Kirchkamp, Tabea Krauss, Orsola Garofalo, Yoan Hermstrüwer, Pascal Langenbach, Natalia Montinari, Michelangelo Rossi, Enrico Santarelli, Enya Turrini, Leon Wardelmann, Caroline Weber, Yien Wu, and the seminar audiences at Latin American and Iberian Law and Economics Association, Aarhus University, the Law, Economics and Regulation Workshop at the University of Paris Panthéon Assas, the ECONtribute LawEcon Workshop at the Max Planck Institute for Research on Collective Goods in Bonn. For making the Main Study possible, we thank Jason Barabas and Herschel Nachlis of the Rockefeller Center at Dartmouth College.

## Author contributions

All the authors contributed equally to this article.

## Competing interests

The authors declare no competing interests.

## Ethical approval

All the experimental procedures were conducted in accordance with the ethical standards for studies with human subjects, and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The study was approved by the Ethics Review Panel of the European Commission as part of the Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 748062 (EC signature date: 30 March 2017). The Pilot Study was conducted in accordance with the ethical standards and guidelines of the LINEEX lab; the Main Study was carried out through an online platform and followed all relevant ethical policies for the recruitment and participation of human subjects.

## Informed consent

Informed consent was obtained electronically from all study participants as part of the survey, prior to the beginning of the experimental tasks, on the same date each individual took part in the study. Because data collection spread across several days, for brevity reasons we report here the date ranges during which informed consent was collected: in the Pilot Study, consent was obtained between November 9 and 14, 2020 in Denmark, and between November 25 and 26, 2020 in Italy; in the Main Study, consent was obtained between August 13 and November 19, 2024 in Italy, between November 23 and December 2, 2024 in Denmark, and on November 7 and December 7, 2024 in the United Kingdom. The consent screen included information on the purpose of the study, data handling, anonymity, payment, and contact details. Participation was entirely voluntary, and participants could only proceed to the experimental tasks by confirming agreement with all the terms shown on the screen through a required click. The full consent form, exactly as it appeared to participants, is included in the experimental instructions provided in the Supplementary Material.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1057/s41599-025-05549-6>.

**Correspondence** and requests for materials should be addressed to Alice Guerra.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025