Why Don't Donors Deduct? Social Norms and the Limits of Tax Incentives[*]

Michael Hilweg-Waldeck[†] & Argun Hild[‡]

Abstract

Many donors leave tax benefits unclaimed even when doing so requires minimal effort and yields meaningful financial rewards. Findings from our representative survey point to confusion about how to deduct donations and to misperceived social norms about the moral appropriateness of doing so as the main drivers of this gap. We study how to tackle these two sources of the deduction gap by providing concise information on how to deduct donations and a one-sentence norm cue in an online experiment (n = 483), a door-to-door field experiment with address-level randomization (n = 6,728), and a radio-based campaign spanning two Austrian federal states. We find that almost all donors deduct when donating through the anonymous online tool. By contrast, during face-to-face fundraising, where social-image concerns are salient, fewer than 1 in 100 donors choose to do so. Across settings, information on how to deduct donations alone leaves deduction behavior unchanged, whereas combining this information with the norm cue increases take-up in the door-to-door setting. Our findings show that financial incentives can falter when clashing with misperceived norms in social settings, unless paired with campaigns that reshape those norms.

[†]Hilweg-Waldeck: Ph.D. candidate, University of Mannheim/ZEW Mannheim, michael.hilweg@uni-mannheim.de
[‡]Hild: Ph.D. candidate, University of Mannheim/ZEW Mannheim, argun.hild@uni-mannheim.de

1 Introduction

Tax incentives are a central tool for policymakers to steer taxpayer behavior toward socially desirable ends, including charitable giving. By lowering the effective cost of donations, governments aim to encourage individuals to direct more of their own money toward causes deemed of public value.[1] Many tax systems implement such incentives by allowing donors to offset part of their charitable contributions through deductions or tax credits. Yet even in systems where these mechanisms are well established and easy to use, many taxpayers still fail to make claim the available benefits.

In 2017, Austria substantially simplified the deduction process of charitable donations by introducing automatic electronic reporting. Previously, donors had to request a receipt from the charity and manually include it in their tax return. Since the reform, charities are required to collect identifying information from deducting donors and transmit it to the tax authority, which automatically prefills the deduction in taxpayers' returns. Yet, despite this simplification, the share of taxpayers deducting donations rose only by 1.8 percentage points (from 19.8% to 21.6%), leaving roughly four-fifths of taxpayers still foregoing the benefit. This modest increase raises a central question: why did take-up remain so limited even after administrative barriers were almost completely removed?

A complementary and equally puzzling pattern emerges when comparing deduction rates across donation channels for one of Austria's largest and most established charities. Among anonymous online donors, 86% claim the tax incentive. In contrast, during a nationwide door-to-door campaign where donors interact directly with local volunteers, only 0.7% claim the deduction, even though donations frequently exceed 50 EUR. This gap of two orders of magnitude underscores the depth of the puzzle and suggests that the determinants of deduction behavior extend beyond filing costs. While observational data reveal the puzzle, they cannot uncover why donors respond so little.

Our collaboration with one of Austria's largest and most established charities offers a unique opportunity to unravel these mechanisms directly in the field. The partnership grants access to a nationwide annual charity event involving more than 800,000 donors and covering all of Austria. This scope enables us to introduce well-defined, randomized treatments within a naturally occurring charitable context and to observe behavioral responses under genuine social conditions. The combination of experimental control and external validity provided by this setting allows us to isolate key drivers of deduction behavior in a way that purely observational analyses cannot.

To understand why so many donors leave tax benefits unclaimed despite negligible filing costs, and why deduction rates differ so sharply between anonymous and socially observable settings, we next examine the potential mechanisms behind this puzzle. We explore two main explanations, both suggested by prior work on taxation and charitable giving. First, donors may lack procedural knowledge about how to claim deductions (Bhargava and Manoli 2015; Chetty et al. 2013; Chetty and Saez 2013; Stantcheva 2021). Second, behavioral barriers such as social-image concerns and misperceived norms may limit the response to monetary incentives even when information is available (Andreoni et al. 2017; Bénabou and Tirole 2006, 2011; Bursztyn and Yang 2022; DellaVigna et al. 2012; Exley 2018).

We begin by examining taxpayers' knowledge and perceptions through a representative survey designed to establish baseline understanding of the deduction system and beliefs about prevailing norms. Two patterns emerge. First, more than 60% of respondents lack

---

[1]See, for example, Republik Österreich (2023) for Austria, Joint Committee on Taxation (2013) for the United States, HM Treasury (2015) for the United Kingdom, République Française (2003) for France, and Deutscher Bundestag (2007) for Germany. These legal documents underscore governments' objective to foster more charitable giving by introducing tax deductibility for donations.

essential procedural knowledge required to deduct. Second, we observe a large gap between respondents' own moral views and their beliefs about others', as over 75% underestimate how many people consider deducting donations morally appropriate. By contrast, we find little evidence that privacy concerns or views about the government's use of non-claimed resources play a major role. These diagnostics align with the literature on tax misperceptions (Stantcheva 2021) and pluralistic ignorance (Bursztyn and Yang 2022) and directly inform our interventions.

Guided by the survey evidence, we designed our main field experiment to test whether the two identified mechanisms—lack of procedural knowledge and misperceived norms—can explain low deduction take-up. The intervention took place during the 2023/24 Carol Singers' charity event, one of Austria's most established nationwide donation campaigns. Each year, local groups visit households across the country, generating more than 800,000 individual donations in face-to-face interactions with charity volunteers. This setting provides a natural environment to study deduction behavior, where social visibility and local norms are salient and where we can link treatment assignment to the charity's registry data on deduction activity.

In collaboration with a set of municipalities comprising roughly 255,000 residents, we implemented a field experiment with two distinct components that together form the core of our analysis. In three municipalities with about 20,000 inhabitants, we manually distributed 10,000 copies of parish journals to implement randomization at the address level. We blocked random assignment by small neighborhood clusters to ensure balance across observable characteristics such as housing type and socioeconomic composition. In the remaining 28 municipalities, we randomized at the municipality level. In both cases, the intervention relied on local parish journals that reach almost all households before the annual event. Each journal contained one of two short paragraphs placed after the main article on the charity event: the *Information* treatment described the four steps required to obtain a tax deduction, while the *Morality* treatment repeated the same text but added a concise norm cue—"88% of participants in a recent representative survey found it morally appropriate to deduct charitable donations." This design allowed us to vary the content of a familiar communication channel of the charity without altering its operations or affecting compliance.

The intervention reached donors in their real decision environment, weeks before the donation moment, allowing it to shape intentions rather than mere recall of information. We fully randomized treatments within the charity's long-established institutional framework, combining high external validity with experimental control. At the municipality level, providing a moral-norm cue increased the number of deductors by about 7.6 per treated ZIP code, corresponding to an effect size of 0.36 standard deviations. At the address level, we do not detect statistically significant effects, which is consistent with the experiment's lower power to detect an effect of this magnitude. Procedural information alone leaves behavior unchanged in both settings, indicating that information is not sufficient when misperceived norms suppress the use of available tax benefits."

To reach a broader population and assess the role of treatment timing and delivery medium, we conducted a second field experiment one year later during the same Carol Singers' event. The intervention consisted of radio advertisements broadcast four times on a single day across two stations in each of two federal states. The broadcast date coincided with the median day of deduction activity as observed in previous years. Based on station reports, the two stations in Upper Austria (*Morality* treatment) reached a daily audience of about 567,000 listeners, while those in Carinthia (*Information* treatment) reached about 558,000 listeners. We again used the charity's registry data on deduction activity to evaluate the effects of the intervention. We find no statistically significant effect of either treatment. Compared with Field Experiment 1, the use of a federal-state broadcast rather than local

media may have made the messages less salient, as listeners likely care more about the views of their immediate community. In addition, legal constraints prevented us from explicitly mentioning the charity or the event in the spots, which may have further reduced treatment strength.

Our evidence so far comes from socially observable contexts, where visibility activates social norms as people's decisions are visible to others. Naturally, it remains unclear whether the same two channels, lack of knowledge and social norms, matter when there is no opportunity to use behavior for social signalling. To assess this, we turn to an online setting that reproduces the deduction decision in a fully unobserved environment, allowing us to examine whether these mechanisms still operate when social visibility, and thus the potential for signalling, is absent. We conducted an online experiment with Austrian and German participants that replicated the key elements of the donation and deduction decisions in a fully anonymous setting. Participants earned money through real effort, were taxed, and could donate part of their earnings to real charities, some of which were eligible for deduction, before deciding whether to deduct. In this baseline condition, 85% of participants chose to deduct, a rate comparable to the charity's own online donation statistics and suggesting that high take-up in anonymous settings is not merely driven by sample composition. Relative to this baseline, we find no significant treatment effects of neither the *Morality* nor the *Information* treatment. The results show that when decisions are private and social signalling is impossible, deduction rates remain high and stable, thus replicating the stark gap between anonymous and observable contexts.

The contrast between field and online evidence suggests that visibility may act as a crucial activator of social norms. Yet these settings differ along other dimensions beyond observability, such as timing, medium, and audience composition. To isolate whether social norms drive the difference, we conducted a laboratory experiment that manipulates only one feature of the decision environment, the visibility of the deduction choice. This feature serves to activate social norms while all other elements are held constant. Participants earned money through real effort, faced income taxation, could donate to a real charity, and then decided whether to deduct. In the *Anonymity* condition, deductions were made privately, whereas in the *Observability* condition, the decision occurred under the gaze of an experimenter or lab assistant standing nearby. Other participants never had access to another's decision, ensuring that only experimenters could observe choices of those assigned to *Observability*. The sample comprises 473 participants. Being observed reduces deduction by about 12 percentage points; however, the estimate is not statistically significant at conventional levels. Open-ended responses describing deduction choices point in the same direction as the quantitative point estimate. Taken together, these findings provide suggestive evidence that varying visibility can activate or deactivate the influence of social norms on the use of tax incentives.

We make two main contributions to research in behavioral public economics and the study of social norms in economic decision-making. First, our large-scale field experiment embedded in Austria's nationwide charity event provides causal evidence on the mechanisms that limit the effectiveness of tax incentives when deduction decisions are socially observable. We show that a concise moral-norm cue substantially increases deduction take-up, whereas procedurally equivalent information alone does not. This complements work in behavioral public economics documenting how informational and cognitive frictions shape responses to tax incentives (Bhargava and Manoli 2015; Chetty 2015; Chetty et al. 2013; Chetty and Saez 2013; Stantcheva 2021) and connects to research emphasizing how social norms and moral perceptions influence economic behavior (Bénabou and Tirole 2006, 2011; Bursztyn and Yang 2022).

Second, by comparing socially observable field contexts with fully anonymous online deci-
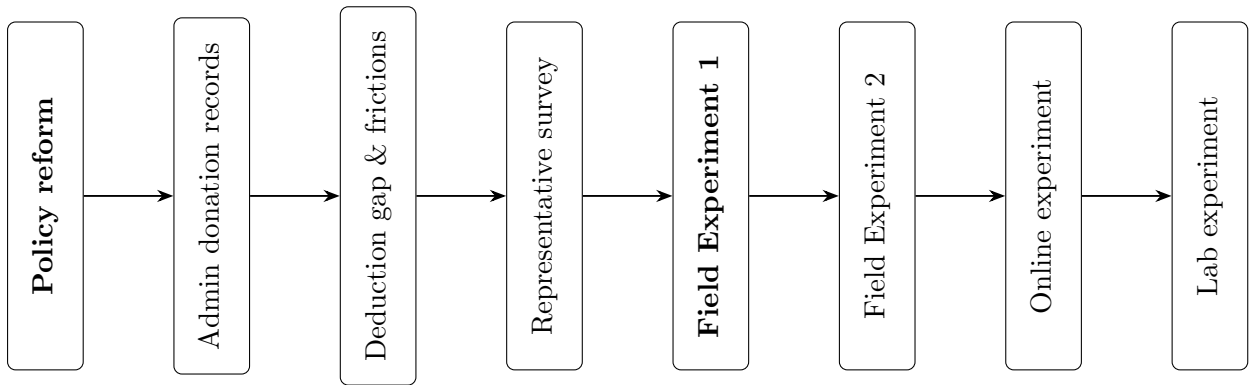
Figure 1: Overview of empirical and experimental components.

sions and a controlled laboratory setting that holds payoffs, information, and tasks constant while varying the degree of observability, we provide causal evidence that visibility activates social norms in the use of available tax benefits. This cross-setting design links the public-economics focus on incentive responses to the norms literature showing how social norms and perceived expectations shape behavior (Allcott 2011; Ariely et al. 2009; Bicchieri 2016; DellaVigna et al. 2012; Ekström 2012; Exley 2018; Lane et al. 2023), providing direct evidence that observability can alter the private calculus behind claiming a benefit.

In addition, we provide a contextual administrative baseline based on tax records from 2013–2019 to quantify the impact of Austria's 2017 automation reform, which removed filing costs for charitable deductions. Despite its ambition, the reform raised deduction take-up by less than two percentage points, with small changes on the intensive margin and modest heterogeneity by age, gender, federal state, and income. This baseline situates our experimental evidence within the national policy environment and shows that administrative simplification alone was insufficient to substantially increase participation.

Taken together, our experiments show that procedural instructions alone leave deduction behavior unchanged. By contrast, moral-norm cues can shift take-up when social norms are active in observable settings. High deduction rates in anonymous online contexts confirm that selection is not driving the puzzle, while the laboratory experiment—where other participants are present in the same room but decisions can be shielded—yields intermediate deduction rates around 53%. This pattern underscores the central role of visibility in activating norms. By combining field, online, and laboratory evidence, we show across complementary settings that misperceived norms, not information frictions alone, explain why financial incentives often fail to deliver their intended effect.

2   A Simple Conceptual Framework

Understanding why most donors do not claim tax deductions even when the process is nearly effortless requires a structured way to think about how individuals weigh private benefits, effort costs, and social considerations. The framework adapted from Bursztyn and Jensen (2017) serves an illustrative purpose: it helps formalizing our thinking and clarifies which mechanisms our experiments are designed to speak to. It is thus not intended for formal model testing, but rather to make transparent how visibility and social norms may condition the behavioral response to financial incentives. In particular, it highlights why we focus on social visibility as a key contextual determinant of behavior.

We model the latent utility from choosing to deduct ($a_i = 1$) rather than not deduct

$(a_i = 0)$ as

$$\tilde{\alpha}_i = B_i - C_i + \nu_i \, \lambda_{ij} \, E_i(\omega_j) \, Pr_{-i}\big(\sigma_i = h \mid a_i\big) + \epsilon_i.$$

Here, B_i captures the private financial benefit of deduction, while C_i denotes the effort cost associated with the procedural step of claiming. The third term represents the social-image component: individual i cares about how others in her reference group j interpret her action. She places weight λ_{ij} on others' opinions and holds beliefs $E_i(\omega_j)$ about the social desirability of being the "high" type, such as a generous, altruistic, or selfless person, within group j. $Pr_{-i}(\sigma_i = h \mid a_i)$ captures others' inference about i's type after observing her action a_i. We extend this framework by introducing an explicit visibility parameter $\nu_i \in [0, 1]$, which scales the relevance of social-image payoffs depending on the extent to which the action a_i acts as a social signal to others. When deduction is anonymous ($\nu_i = 0$), it does not convey a signal and social image concerns vanish; when the action is observable ($\nu_i > 0$), its signaling potential increases, and thus its reputational relevance.

This framework helps interpret the empirical facts discussed in the introduction: the 2017 reform removed administrative frictions but produced little behavioral change, while deduction rates differ sharply between anonymous and visible settings. Such patterns suggest that variation in the visibility of the decision and in the perceived social meaning of deduction may be central to understanding donors' behavior. Our experimental evidence builds on this idea by systematically varying how observable the action is and how it may be interpreted by others, with particular emphasis on socially visible environments.

3 Empirical Evidence on Deduction Behavior and Underlying Mechanisms

3.1 The Carol Singers' Charity Drive: A Door-to-Door Tradition

To study why donors often fail to claim available tax deductions, we need a setting that combines three features: (i) broad population coverage to capture the behavior of a diverse donor base, (ii) a high degree of social interaction so that social norms and image concerns are activated, and (iii) access to reliable administrative records that allow us to track deduction behavior. The Austrian Carol Singers' Charity Drive (*Dreikönigsaktion*) offers exactly such a setting. As one of Austria's ten largest NGOs by donation revenue, it operates a standardized nationwide campaign with deep local roots through the Catholic parish structure.[2] Founded in 1954, the organization has become a widely recognized part of Austria's charitable landscape, ensuring both visibility and credibility. Most importantly, its door-to-door fundraising format creates direct, face-to-face interactions between donors and local volunteers, making it an ideal environment to examine how observability shapes tax deduction behavior.

The event takes place annually between December 26 and January 6, when groups of children and adults dressed as kings and queens visit households, sing songs, and collect donations (Figure 2). Over time, the drive has become one of Austria's most visible and anticipated charitable traditions, with households across the country expecting the singers' annual visit.

---

[2]In Austria, parishes are local Catholic communities that typically overlap geographically with municipalities and serve as the organizational base for the campaign. Outside the major cities, parish and municipal boundaries largely coincide. Carol Singing covers all households independent of religious affiliation and is not perceived as a religious activity, making the parish effectively equivalent to a municipality for our purposes.

Figure 2: Group of Austrian carol singers (© Dreikönigsaktion)

As part of our collaboration, we obtained access to detailed administrative data on both donations and deductions and were granted permission to embed experimental treatments during the annual event.

3.2 Large Differences in Deduction Behavior Across Settings

Studying deduction behavior in charitable giving is challenging because large-scale data from different donation channels are rarely available. In particular, field data from in-person interactions are typically limited to deductors only, making it difficult to assess how many donors forgo the incentive. Our collaboration with the Carol Singers' Charity Drive provides a rare opportunity to overcome this challenge. We combine two complementary data sources that allow us to compare deduction behavior across highly observable and anonymous contexts.

First, we use administrative records from the door-to-door campaign, where donations are made in face-to-face interactions with local volunteers. Since 2017, Austrian NGOs eligible for donation deductibility have been required to collect and transmit donor information to the tax authority. During the drive, this means that deductors provide their personal details (name, date of birth, address, amount donated) to the accompanying adult, while non-deductors simply hand over their donation without any paperwork. To capture the behavior of this otherwise invisible group, we organized a complementary data-collection effort: in five municipalities, accompanying adults recorded anonymous tick marks for each donation received. This yielded around 4,500 ticks from 10,500 households, corresponding to a donation volume of 78,000 EUR (see the tick-list template reproduced as Figure 37 in Online Appendix B.1). These tick lists allow us to construct the denominator needed to calculate the deduction rates in the field.

Second, we analyze administrative data from the charity's online donation platform,[3] where donors face the same incentive but make their decision without social interaction. Online donations are heavily concentrated during the same period as the drive (see Figure 9 in Appendix A.1). Unlike the field data, the online platform records both deductors and non-deductors, including information on personal characteristics and donation amounts.

Taken together, these two sources provide a unique window into deduction behavior across contrasting social environments. Consistent with anecdotal evidence and prior behavioral research, deduction rates differ sharply across settings. Almost nine out of ten online donors claim a deduction, whereas in the field setting only about one in two hundred do so. Table 1 summarizes these patterns, reporting both average donation amounts and approximate donor counts. For the field, we estimate a total of about 800,000 cash donors nationwide per year, while online figures are drawn directly from the charity's administrative records.

	Share of deductors	# Donors	Average donation (deductors)	Average donation (non-deductors)
Field	0.66%	∼800,000	71.72 EUR	17.32 EUR
Online	86%	19,042	54.31 EUR	36.46 EUR

Table 1: Deduction propensities and average donations by deduction status across settings.

3.3 Uncovering Mechanisms: Norm Perceptions and Information

To probe why such stark differences emerge across settings, we complement these administrative data with a representative survey on taxpayers' knowledge and perceptions. We conducted an online survey of Austrian taxpayers (n=314) in April 2023 via Bilendi, an online research panel provider that offers quota-based national representativeness in terms of gender, age, and education. The survey included several components. First, participants reported whether they had donated in the past and, conditional on donating, whether they had claimed a deduction. Second, we assessed their procedural knowledge of the tax deduction process by asking about four key steps: providing personal data when donating, submission of information by the charity to the financial authorities, automatic consideration of the deduction in the annual tax declaration, and automatic transfer of the reimbursement. Third, we elicited whether they personally considered it appropriate to deduct and beliefs about the share of others finding it appropriate to do so. Finally, we asked for open-ended reasons in favor of and against deducting.

The survey also contained a vignette designed to probe perceptions of generosity in the presence of tax deductions. Respondents were told about "Lukas," who considers either (a) donating 30 EUR without deducting, or (b) donating 35 EUR and claiming an 8 EUR reimbursement via deduction. Participants indicated which option they regarded as more generous and provided a short justification. Results from this vignette are presented in Appendix A.2.

The main results reveal two principal dimensions. First, substantial informational gaps exist, with 64% of respondents failing to correctly identify at least one step in the deduction process. Second, social image concerns are evident as we observe a 24 percentage point gap between respondents' own views of moral appropriateness and their beliefs about how others view deduction.

As additional descriptive patterns, we find that moral attitudes toward deducting strongly predict behavior. Respondents who considered tax-deducting donations "appropriate" were

---

[3]See the Dreikönigsaktion's online donation portal at https://www.dka.at/spenden/online.
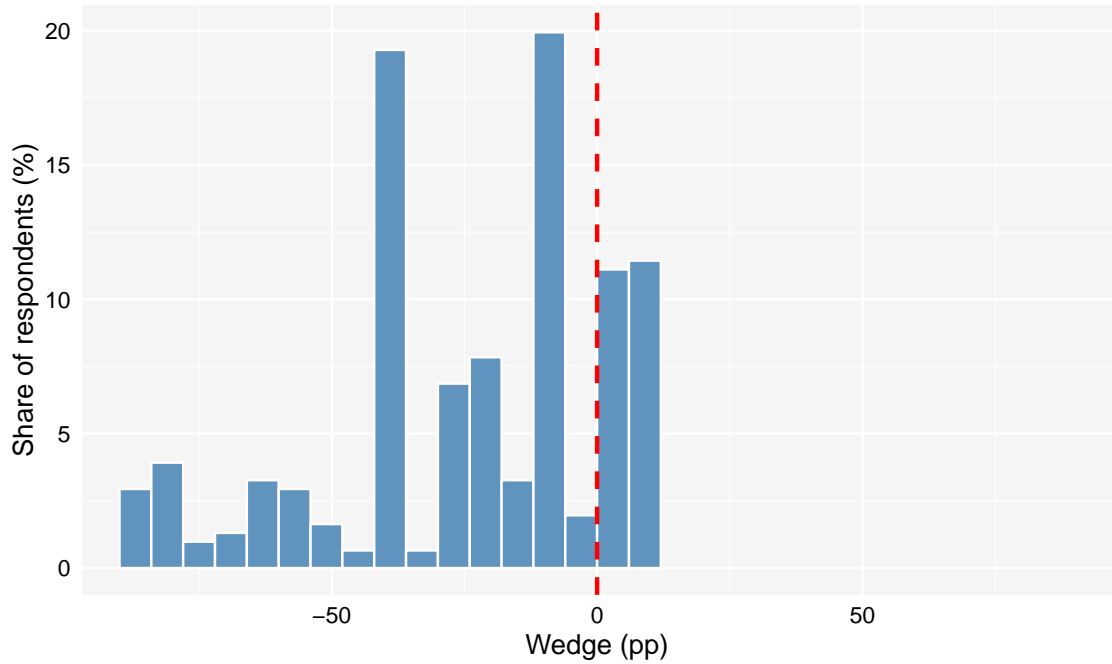
Figure 3: Distribution of perceived social support for deducting donations

The figure shows the deviation between respondents' beliefs about others' views and the actual average view in the sample. Negative values indicate underestimation, positive values overestimation. The dashed red line marks the actual average view.

substantially more likely to deduct themselves, whereas those who judged it "inappropriate" were predominantly non-donors or non-deductors (see Appendix A.2, Figure 15). Likewise, perceptions of procedural effort align with experience: deductors reported lower expected time costs than non-donors or non-deductors, consistent with beliefs being shaped by prior exposure and misperceptions (Appendix A.2, Figure 17).

Figure 3 visualizes the distribution of wedges, defined as the difference between each respondent's belief about others' views and the actual average view. The bulk of the mass lies to the left of zero, showing that more than 75% of respondents underestimate support for deducting. This systematic misperception suggests that norm-related frictions, not just lack of procedural knowledge, are key to understanding low take-up of tax incentives.

4 Social Visibility and Deduction Behavior

The survey evidence reveals two clear frictions: missing procedural knowledge and systematically misperceived social norms. To test whether these barriers causally explain low take-up, we embedded a field experiment directly in the 2023/24 Carol Singers' charity event. This setting is uniquely suited to our research question since donation and deduction decisions occur in face-to-face interactions, where social norms are highly salient and behavior is highly observable, and administrative deduction records can be linked to treatment assignment.

4.1 Design

We delivered our interventions through the final annual issue of local parish journals, which reach virtually all households in parishes and are directly tied to the charity event. The journals carried a feature article marking the 70th anniversary of the event. Depending

on treatment assignment, this article was followed by one of two treatment messages. The *Information* treatment outlined the four procedural steps required to obtain a deduction: (i) providing personal details when donating, (ii) the charity transmitting this information to the tax authority, (iii) automatic pre-filling of the deduction in the annual tax return, and (iv) the subsequent refund being transferred to the donor's account. The *Morality* treatment combined these instructions with a concise norm cue: "88% of participants in a recent representative survey found it morally appropriate to deduct charitable donations." The exact wording of both treatments is provided in Appendix A.4. Interpreted through the lens of the conceptual framework, the two messages operate through distinct channels. The *Information* treatment can only lower perceived effort costs if donors initially overestimate the procedural steps required to claim a deduction (C_i). The *Morality* treatment, by contrast, connects to the social-image component by shaping how individuals believe others view deduction in moral terms. In this sense, it affects the inference term $\Pr_{-i}(\sigma_i = h \mid a_i)$: it helps donors recognize that deduction is generally seen as consistent with being generous or selfless.

The intervention covered 30 parishes or parish unions with a combined population of roughly 235,000 residents. Parish unions are administrative clusters of several parishes and function similarly to municipality unions. In 29 units, randomization occurred at the parish level, with all households in a unit receiving the same journal version. In one parish union, we implemented address-level randomization by manually delivering nearly 10,000 copies over ten days. We blocked assignment by carol-singing routes to ensure comparability across neighborhoods. Each route consists of a small number of adjacent streets, and households within a route are typically similar in socioeconomic status. Blocking at the route level therefore ensures that treatment assignment was balanced across key observables. This fine level of randomization was only possible in one parish union since the delivery was carried out entirely by the two researchers, which put a strict cap on the number of addresses that could be covered in this way.

Parishes were recruited through referrals from the charity's state-level coordinators, personal contacts, and targeted cold-calling. Treatment assignment was guided by three balancing criteria: geographic dispersion across federal states, population size, and the degree of urbanization. Achieving balance required navigating rolling recruitment and parish-specific publication deadlines. Some parishes had to be assigned before others were even recruited, which limited perfect ex-ante balance but allowed all units to be included on time. Figure 4 displays the geographic distribution of treatment units. While the map shows an uneven spread across Austria, identification relies on randomized assignment at the parish level and on difference-in-differences comparisons of changes in the number of deductors between treated and untreated parishes.

4.2 Results

We evaluate treatment effects using both the parish-level assignments and the one parish union where address-level randomization was implemented. For the parish-level analysis, we observe deduction data for the December 2022/January 2023 event (pre-treatment) and the December 2023/January 2024 event (post-treatment). This allows a difference-in-differences design that compares treated and untreated units before and after the intervention, while at the same time correcting for any location-specific factors that are constant over time. The outcome variable is the number of deductors in a given parish–year or ZIP–year cell. Because the parish-level analysis was not part of our preregistration, we report all plausible
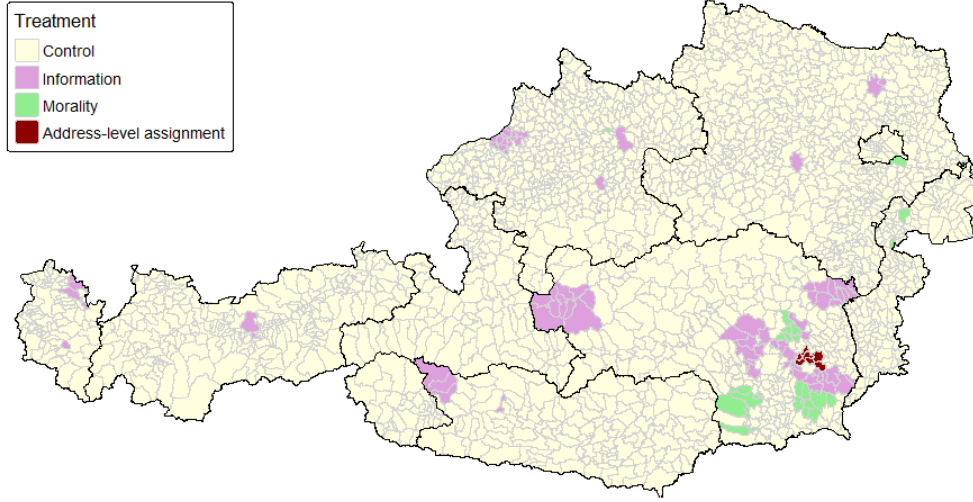
Figure 4: Treatment area of parish-journal intervention.

specifications for transparency.[4]

The most precise specification is the active ZIP sample, defined as ZIPs with at least one deduction in either year. We first analyze the pooled version of both treatments to assess whether the intervention had any effect. Our staggered treatment approach enables us to do so because both arms contain the same procedural information, with Morality adding a short norm cue on top. In the active ZIP specification, the pooled treatment increases the number of deductors by 2.38 per treated ZIP ($p = 0.034$), equivalent to 0.125 SD using the pre-treatment distribution of deductors (Figure 5, Table 3). We then decompose the pooled effect by arm. The *Morality* treatment increases deductors by 7.6 per treated ZIP ($p < 0.01$), or 0.36 SD, whereas the *Information* treatment shows no measurable impact (Figure 6, Table 2). Combining the insights of the pooled and the treatment-split analysis clarifies that there is an overall treatment effect and that it is driven by the norm cue rather than by procedural instructions alone.

At the address level, where assignment occurred within one parish union, the study is underpowered to detect effects of the magnitude we estimate at the ZIP level. Baseline take-up was only about 0.66% of households, so the *Morality* arm contained roughly 12–13 deductors before the intervention. An effect size comparable to the ZIP-level estimate would translate into about 5 additional deductors in this arm. Yet with about 2,200 addresses per arm, the minimum detectable effect size at 80% power ($\alpha = 0.05$) is roughly +0.68 percentage points, or about 15 deductors. The expected effect therefore falls well below the study's detection threshold, which explains the null finding in the address-level data. By contrast, the ZIP-level analysis covers nearly 4,400 units observed over two years, with 1,563 active ZIP–year observations in the preferred sample. The outcome variable, the number of deductors per unit, is more continuous and less sparse than binary household-level take-up, which provides substantially greater statistical power. This explains why moderate effects that are undetectable at the address level can nevertheless be estimated with precision in the ZIP-level analysis.

We also test for effects on the intensive margin by analyzing the average donation amount among deductors. Across all specifications, we find no statistically significant changes (Appendix A.4). This pattern suggests that newly converted deductors adjusted their contributions upward to the higher donation levels typical of established deductors, leaving the

---

[4]The study was preregistered with a pre-analysis plan focused on the address-level randomization, available at https://osf.io/5ew3j.
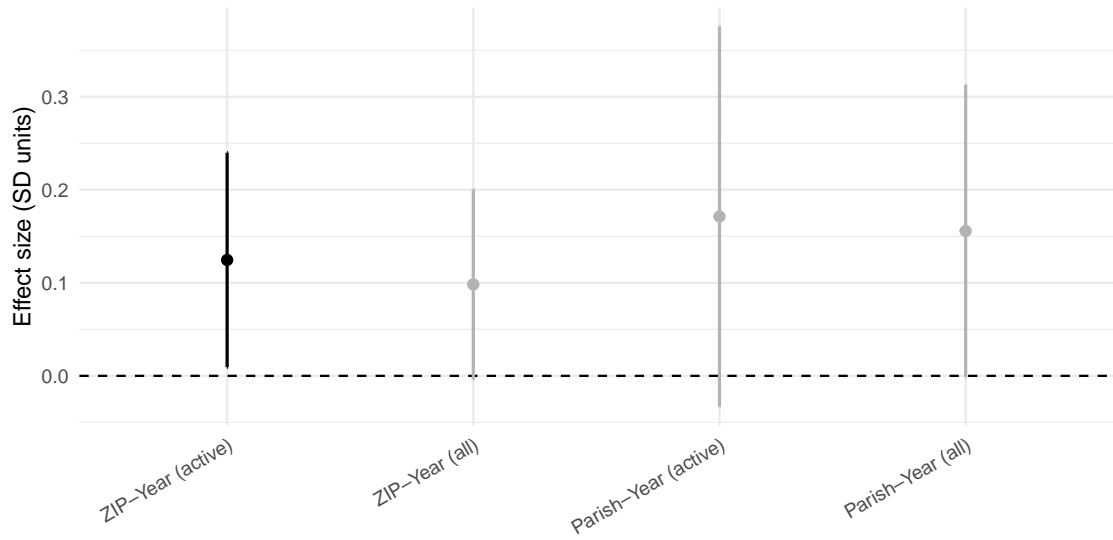
Figure 5: Coefficient plot of pooled treatment effects from Field Experiment 1.

The figure displays point estimates and 95% confidence intervals for pooled treatment across all four specifications (ZIP–Year active, ZIP–Year all, Parish–Year active, Parish–Year all). The coefficients are expressed in standard deviation units of the pre-treatment distribution of deductors. Estimates for active ZIP–Year are highlighted in black; all others are shown in grey.

Figure 6: Coefficient plot of treatment effects from Field Experiment 1.

The figure displays point estimates and 95% confidence intervals for both the *Information* and *Morality* treatments across all four specifications (ZIP–Year active, ZIP–Year all, Parish–Year active, Parish–Year all). The coefficients are expressed in standard deviation units of the pre-treatment distribution of deductors. Estimates for active ZIP–Year are highlighted in black; all others are shown in grey.
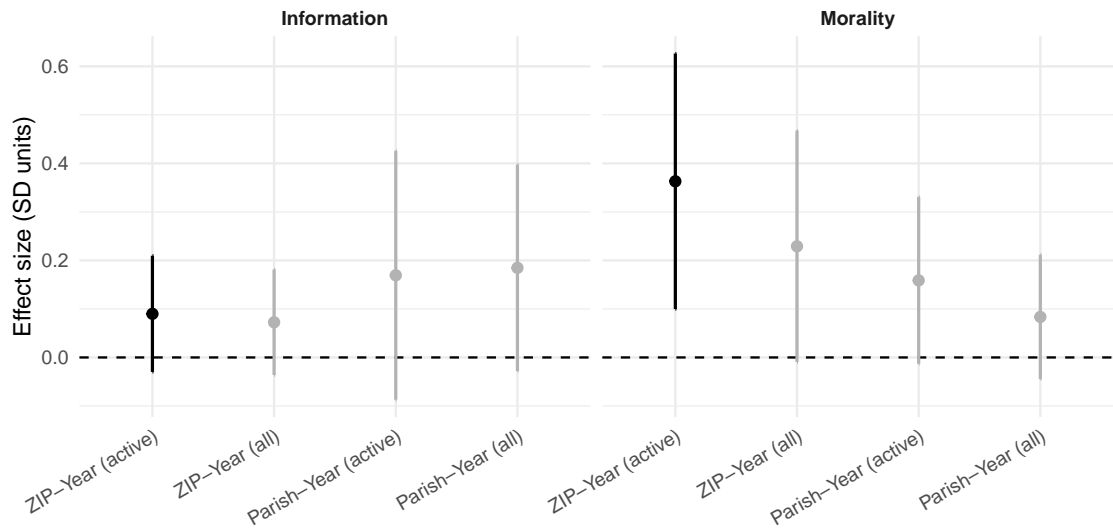
average donation among deductors unchanged.

	ZIP-Year (active)	ZIP-Year (all)	Parish-Year (active)	Parish-Year (all)
Treated	2.378**	1.875*	4.767	4.028*
	(1.123)	(0.994)	(2.901)	(2.072)
	(0.034)	(0.059)	(0.101)	(0.052)
Num. Obs.	1640	4472	1587	4318
R^2	0.954	0.923	0.954	0.924
SE	ZIP	ZIP	Parish	Parish
FE	Unit	Unit	Unit	Unit
	+	+	+	+
	Event-Year	Event-Year	Event-Year	Event-Year

Table 2: Difference-in-differences estimates of pooled treatment effects on the number of deductors.

The outcome variable is the number of tax-deducting donors per unit–year (ZIP or parish). "Active" samples are restricted to units with at least one deduction in either the pre- (Dec 2022/Jan 2023) or post-period (Dec 2023/Jan 2024). Point estimates represent the absolute change in the number of deductors per treated unit relative to controls. Standard errors (clustered on unit level) and p-values in parentheses. All models include two-way fixed effects (unit and event-year). * p $<$0.1, ** p $<$0.05, *** p $<$0.01.

	ZIP–Year (active)	ZIP–Year (all)	Parish–Year (active)	Parish–Year (all)
Information	1.699	1.369	4.892	5.070*
	(1.153)	(1.047)	(3.778)	(2.978)
	(0.141)	(0.191)	(0.196)	(0.089)
Morality	7.642***	4.823*	4.392*	1.945
	(2.828)	(2.561)	(2.425)	(1.521)
	(0.007)	(0.060)	(0.070)	(0.201)
Num.Obs.	1640	4472	1587	4318
R2	0.954	0.923	0.954	0.924
SE	ZIP	ZIP	Parish	Parish
FE	Unit	Unit	Unit	Unit
	+	+	+	+
	Event-Year	Event-Year	Event-Year	Event-Year

Table 3: Difference-in-differences estimates of treatment effects on the number of deductors.

The outcome variable is the number of tax-deducting donors per unit–year (ZIP or parish). "Active" samples are restricted to units with at least one deduction in either the pre- (Dec 2022/Jan 2023) or post-period (Dec 2023/Jan 2024). Point estimates represent the absolute change in the number of deductors per treated unit relative to controls. Standard errors (clustered on unit level) and p-values in parentheses. All models include two-way fixed effects (unit and event-year). * p $<$0.1, ** p $<$0.05, *** p $<$0.01.

Taken together, the evidence from Field Experiment 1 shows that providing a moral-norm cue increases deduction take-up, while procedural instructions alone have no effect. The results highlight that social-image considerations, when activated in a natural high-visibility setting, can meaningfully shift behavior even in a context with long-standing and entrenched donation habits. This contribution is twofold. First, we provide causal evidence on how social norms affect the use of tax incentives in real-world giving. Second, we demonstrate a scalable, low-cost way of correcting misperceived norms to increase take-up, while showing that purely informational interventions—i.e., efforts that only inform the public about the availability and mechanics of tax deductibility—are ineffective in such settings.

5 Scaling and Timing of Norm Activation

The first field experiment shows that norm cues can shift behavior in socially embedded settings. But the intervention's success raises two open questions: can such cues be scaled to larger audiences, and does their dissemination weeks before data collection lead to diluted effects? To address these points, we implemented a second field experiment that delivered treatment via radio broadcasts during the donation period itself.

5.1 Design

We implemented the intervention on January 4, 2025, during the peak of the Carol Singers' charity drive. While the drive formally spans December 26 to January 6, past data show
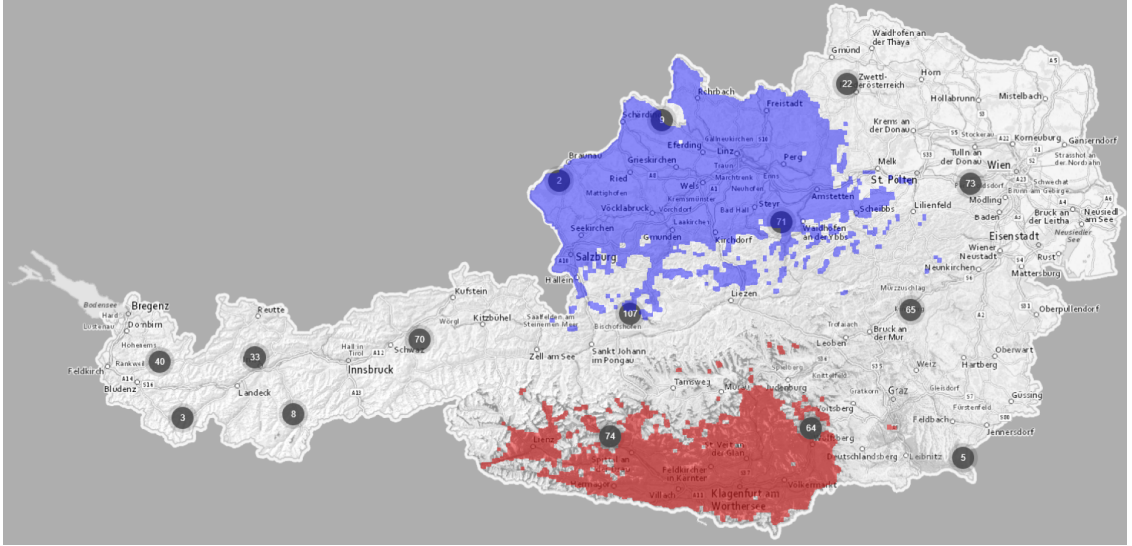
Figure 7: Treatment area of radio-ad intervention.

The figure shows the radio broadcast coverage areas for the *Information* treatment (red) and the *Morality* treatment (blue).

that in Upper Austria and Carinthia nearly all deduction activity occurs between January 2 and 6, with January 4 as the median day. We therefore selected this date for broadcasting. Each of the two states was assigned one treatment: the *Morality* message in Upper Austria and the *Information* message in Carinthia. To ensure broad demographic coverage, we recruited two radio stations: a state-level public broadcaster (ORF) with the largest daily audience and a private station (Kronehit) targeting younger listeners. In both states, the spot was aired four times on January 4 on each station.[5] Because Austrian broadcasting regulations require that radio spots disclose their institutional source, we collaborated with the Vienna University of Economics and Business to include its name as the sender. Referring to an Austrian university ensured credibility and avoided potential suspicion that might have arisen from naming a German institution. Interpreted through the conceptual framework, this second intervention targets the same social-image mechanism as the parish-journal study, but through a different communication channel and timing. Both settings involve the same highly visible donation context, where deduction decisions can signal moral type to others ($\nu_i > 0$). However, by delivering the message via radio and during the donation period itself, this experiment alters how and when individuals may update their beliefs about how deduction is viewed by others, the inference term $\Pr_{-i}(\sigma_i = h \mid a_i)$, relative to the earlier, print-based intervention.

Alongside the radio campaign, we created a dedicated website to provide additional information and tools for potential donors. The site explained the deduction process, featured a searchable registry of all eligible Austrian charities (linked to their official websites), and included a two-way calculator allowing users to estimate either the refund they would receive by deducting or the gross amount they could donate for a given net cost. The website URL was mentioned in both radio spots as the main reference for further information. The site was live from mid-December 2024 through mid-January 2025, used an Austrian domain, and was indexed through Google services to ensure visibility in search results. Online Appendix B.5 provides screenshots of the website's pages.

---

[5]Daily audiences are approximately 567,000 in Upper Austria and 558,000 in Carinthia.

5.2 Empirical strategy

Identification builds on two complementary sources of variation. First, within treated states, the timing of carol-singing routes across January 2–6 is determined primarily by the availability of volunteer groups rather than household or neighborhood characteristics, providing quasi-random assignment of households to pre- versus post-broadcast periods. This allows for a difference-in-differences (DiD) design comparing deduction behavior in January 2–3 (pre) versus January 4–6 (post) within each state. Second, exposure drops discontinuously at the predicted edge of the radio signal. This supports a geographic regression discontinuity design (RDD) that compares municipalities just inside and just outside the coverage area.

Difference-in-Differences (DiD). Our first strategy exploits within–state timing around the single broadcast day, using two years of data (2024 and 2025). The unit of observation is the municipality–day, and the analysis is run separately for each treated state (Upper Austria for *Morality*, Carinthia for *Information*). We define January 2–3 as the pre–period and January 4–6 as the post–period in both years, with January 4 marking the broadcast date in 2025. Identification comes from the quasi-random timing of carol-singing routes: within municipalities, volunteer groups schedule visits based on their own availability and route logistics rather than household or neighborhood characteristics. This generates as-if random assignment of households to pre– vs. post–broadcast days.

Our main specification is a Poisson difference-in-differences model,

$$Y_{mty} = \alpha_m + \delta_y + \gamma \, \mathrm{Post}_t + \beta \left(\mathrm{Post}_t \times \mathrm{Year}_{2025} \right) + \varepsilon_{mty},$$

where Y_{mty} is the count of deductors in municipality m on day t of year y, α_m are municipality fixed effects, δ_y year dummies, and β captures the treatment effect of the 2025 broadcast by comparing the pre–post change in 2025 to the same pre–post change in 2024. Standard errors are clustered at the day level.

As extensions, we estimate (i) an event-time specification probing pre-trends, (ii) a two-way fixed-effects model absorbing all municipality and date heterogeneity, and (iii) a triple-difference design that further compares treated states to the seven non-treated federal states.

Regression Discontinuity (RDD). Complementing our DiD analysis, we implement a geographic RDD that leverages the discontinuous change in exposure at the predicted broadcast border. Official signal propagation data from the Austrian broadcast control agency are geo-referenced and linked to municipality boundaries to classify exposure.[6] The running variable is the signed distance (in kilometers) from the municipality centroid to the broadcast border. We estimate local linear regressions on either side of the cutoff with triangular kernel weights. Our main specification uses a ± 20 km bandwidth, and we report additional estimates for ± 10 km and ± 30 km as robustness checks. We also implement a donut RDD excluding municipalities within 1 km of the border. Identification requires continuity of potential outcomes at the border; we assess this through smoothness tests on pre-treatment deduction rates and observables, density diagnostics, and the donut exclusion. The coefficient of interest is the discontinuity in deduction rates at the cutoff, which under these assumptions captures the causal effect of broadcast exposure. Taken together, the DiD and RDD strategies exploit complementary sources of quasi-experimental variation providing a robust framework to evaluate the causal impact of radio exposure on deduction behavior.

---

[6]https://senderkataster.rtr.at

5.3 Results

Because individual-level data are available only for deducting donors, we deviate from the preregistration.[7]

For Upper Austria, placebo checks support the parallel trends assumption and thereby allow causal interpretation of the DiD estimates. Consistent with the RDD, these results show no detectable effect of the *Morality* treatment (Table 4). For Carinthia, the placebo test suggests a pre-trend, invalidating the DiD design for causal identification. We therefore rely on RDD estimates as the primary evidence, while reporting the DiD output together with placebo tests in Appendix A.5. Across different bandwidths, the RDD estimates consistently indicate a null effect of the *Information* treatment (Table 5). Further robustness checks and analyses of dynamic effects are provided in Appendix A.5.

Two factors likely contribute jointly to this outcome. First, regulatory restrictions on radio-based public service announcements prohibited explicit mention of both the charity name and the event, making it harder for listeners to connect the information to the actual donation opportunity. Second, discussions with our media agency partner and the participating radio stations suggested that a minimum broadcast period of about two weeks would normally be required to achieve sufficient audience reach and recall, whereas budget constraints limited us to a single day of broadcasts. Beyond these structural constraints, it is also possible that information about social norms carries greater weight when communicated through local-level outlets such as parish journals, where it is more naturally perceived as a signal about one's immediate community, rather than through state-wide radio campaigns that convey norms of a more anonymous mass audience. This interpretation is necessarily tentative, but it highlights that the perceived social context of a message may shape its effectiveness.

	# Deductors (percent change)
Upper Austria (Morality)	−33.7
	(25.5)
	(0.284)
Num.Obs.	10
Pseudo R2	0.295
SE	date
FE	state + date

Standard errors and p-values in parentheses. * p <0.1, ** p <0.05, *** p <0.01

Table 4: Poisson DiD estimates of *Morality* treatment effects on daily deductors

---

[7]The preregistration is available at https://osf.io/x45kv.

	Carinthia (Information)	Upper Austria (Morality)
# Deductors (percent change)	-38.4	6.4
	(20.7)	(28.2)
	(0.124)	(0.807)
Num. Obs.	n_L/n_R: 18 / 24	n_L/n_R: 29 / 178

Table 5: RDD estimates of the change in the number of deductors (2025)

Standard errors and p-values in parentheses. n_L and n_R denote the number of observations to the left (outside treatment area) and right (inside treatment area) of the cutoff with 20km bandwidth. * p <0.1, ** p <0.05, *** p <0.01.

Finally, website engagement during the campaign period was very low, yielding insufficient data for formal analysis.

6 Anonymity and the Absence of Social Signals

To benchmark deduction take-up under conditions where social-image concerns are absent, we implemented an online experiment. Our results so far suggest that social observability strongly affects deductibility: while the majority of donors deduct in the anonymous online setting, very few do so in the field. In the latter context, our field experiments indicate that moral-norm cues can shift behavior, likely because social-image concerns hold donors back. What remains unclear is whether such interventions matter at all when anonymity rules out observability pressures. The online experiment therefore provides a clean test of whether information or norm cues retain any power in the absence of social-image concerns. Interpreted through the conceptual framework, this experiment corresponds to the limiting case of full anonymity, where visibility (ν_i) equals zero and deduction carries no social signal. In this setting, others cannot observe the decision to deduct, so the social-image component, captured by $\lambda_{ij} E_i(\omega_j) \Pr_{-i}(\sigma_i = h \mid a_i)$, is effectively switched off.

6.1 Design

We conducted the study on Prolific in October 2023 (n=483) using `oTree` (Chen et al. 2016), recruiting Austrian and German participants.[8] Participants first completed two real-effort encoding tasks to earn money. A flat 30% tax was levied on earnings and transferred into a shared fiscal pool, creating an externality without requiring interaction.

Afterwards, participants could decide whether to donate part of their earnings. Donors could allocate funds across nine real Austrian NGOs. Exactly three were tax-deductible, while six were not. All organizations were presented under neutral labels, but participants could click to view short descriptions including whether the charity was eligible. Actual money was transferred to the selected NGOs after the experiment.

Conditional on donating to at least one deductible NGO, participants were asked whether they wished to deduct. Doing so required them to enter their Prolific ID, mimicking the real-world step of providing personal information. Refunds were financed from the shared fiscal pool, ensuring real monetary consequences. Participants were randomly assigned to one of three conditions. In the *Control* group, deductibility was described minimally: donors

---

[8] See https://osf.io/mruz4 for the preregistration and pre-analysis plan.

were told that if they opted to deduct, they would have to complete the required steps and receive back 30% of their donation. The *Information* treatment added procedural detail by explaining where eligibility information could be found and an explanation that entering the Prolific ID sufficed to complete the process. The *Morality* treatment included all of these elements as well. In addition, it stressed that deductibility enables donors to support those in need more strongly and noted that in a previous study 88% of participants considered deducting donations morally appropriate.

6.2 Results

Neither the *Information* treatment nor the *Morality* treatment produced a statistically detectable effect on deduction propensity. This pattern reflects a ceiling effect: because only donors to at least one deductible charity face the deduction decision (Control n=94, Information n=92, Morality n=101) and take-up in the *Control* group is 85%, the minimum detectable effect size at 80% power is about 11.6–11.8 percentage points per treatment–control comparison.[9]

Treatment messages also left donation behavior unchanged: both the incidence of giving and the allocation between deductible and non-deductible NGOs were unaffected (see Appendix A.3 Figures 18 and 19). Thus, in an anonymous environment, deduction was already close to universal, leaving limited scope for either procedural information or norm cues to shift behavior.

7 Causal Evidence on Visibility and Norm Activation

Deduction take-up varies sharply across contexts: in anonymous online settings, the vast majority of donors deduct, whereas in face-to-face field settings, very few do so. This striking gap suggests that social visibility may suppress deduction, rather than differences in the charitable decision itself. Field interventions, however, cannot cleanly disentangle observability from other features of in-person interaction, since the legal framework imposes strict constraints on how deduction can be implemented.[10] To isolate the role of visibility, we therefore implemented a laboratory experiment that directly manipulated whether deduction decisions were observable to others. This setting allows us to causally test whether making the deduction choice visible activates social-image concerns and thereby shifts behavior. In the language of the conceptual framework, the lab design provides a direct way to vary the visibility parameter (ν_i) while holding financial incentives and procedural costs constant. By doing so, it isolates how the presence of an observer can activate or mute the social-image component, represented by $\lambda_{ij} E_i(\omega_j) \Pr_{-i}(\sigma_i = h \mid a_i)$, that may otherwise discourage deduction in face-to-face settings.

7.1 Design and Treatments

The structure of the lab experiment closely followed the online study while adding experimental control over observability. Participants completed two real-effort tasks to earn income, which was taxed at a flat rate of 30 percent. They could donate part of their earnings to a real charity, Dreikönigsaktion, the same partner as in the field studies. Donations were transferred for real, and deductions were reimbursed at the marginal rate from a shared

---

[9] ≈12.5 pp with a Bonferroni adjustment.

[10] Austrian charities are legally required to verify that the amount claimed for deduction matches the donation actually received. This prevents experimental variation in whether deduction choices are visible, as any attempt to conceal information would risk enabling tax fraud.

tax pool. This setup created realistic stakes and preserved the trade-off between effort and refund.

The design combined two treatment dimensions in a 2×2 structure. To vary observability, participants moved to a separate terminal after completing the donation stage. In the *Anonymity* condition they made their deduction choice alone, while in the *Observability* condition the experimenter escorted them to the terminal, explained that further tasks would be completed there, and remained standing behind them with full view of the screen. Other participants never saw anyone's decision, which ensured that the manipulation focused exclusively on the presence of a neutral observer. To vary exposure to norms, participants in the *Morality* condition saw a short message on their original terminal immediately before moving to the deduction station: *"In a recent representative survey, 88% of respondents said it is morally appropriate to deduct charitable donations."* The experimenter was blind to this assignment. The experiment was implemented in `oTree` (Chen et al. 2016) and coordinated through ORSEE (Greiner 2015) across two Vienna economics laboratories.

7.2 Results

Following our preregistration and pre-analysis plan, we first compared deduction rates between *Anonymity* and *Observability* using a one-sided Boschloo test.[11] We do not reject the null ($p = 0.672$). Likewise, we find no significant effect of the moral message, either among anonymously deducting donors ($p = 0.756$) or among donors whose decision was observed ($p = 0.857$). Because treatment was assigned before the donation decision, the resulting donor samples in the four treatment cells were not fully balanced in demographics even though no treatment content was revealed pre-donation. Summary statistics show modest differences in age and gender (Appendix A.6, Table 19). To address potential covariate imbalance and improve precision, we complement the preregistered tests with a Lin-style OLS linear probability model (Lin 2013) that retains the Observability × Moral-message interaction and adds full interactions between each treatment indicator and the covariates (gender, age, education), while including session fixed effects and HC1 standard errors clustered at the session level. In this specification, deduction rates were 13.9 percentage points lower under *Observability* compared to *Anonymity*, but the effect is imprecisely estimated ($p = 0.196$). The moral message produced small and statistically insignificant effects in both visibility conditions. Figure 8a displays the average marginal effects for *Observability* and *Moral message* as well as the interacted versions for *Observability*.

A post-hoc calculation shows that the minimum detectable effect given the effective sample size we had with 80 percent power and $\alpha = 0.05$ is roughly 23.6 percentage points for a one-sided test (26.6 pp two-sided). This threshold exceeds the observed effect and explains why statistical significance is not reached. Deduction rates in the lab averaged about 59 percent under *Anonymity* and 47 percent under *Observability*, with a mean of roughly 53 percent across both conditions. These levels sit between the high take-up observed in the online experiment and our charity partner's online data (around 86%) and the very low levels in the field (around 0.7%). The contrast in the lab was inevitably weaker as the *Anonymity* condition was less private than at-home giving, since participants remained in a lab environment, and the *Observability* condition involved a neutral lab assistant rather than charity-affiliated volunteers. Together with lower donation incidence in the lab (about 39% compared to roughly 50% online), these features limited power. We therefore view the lab results as a consistency check. While the estimates do not reject the null at conventional levels, they are directionally aligned with the broader evidence that visibility and thus active social norm concerns discourage deduction.

---

[11]See https://osf.io/uc7jw for the preregistration and pre-analysis plan.

To further probe inference, we complement conventional cluster-robust standard errors with a randomization inference (RI) procedure. RI tests the sharp null of no treatment effect by repeatedly reassigning the realized treatment labels within sessions and recalculating the test statistics. This approach exploits the actual randomization scheme, thereby providing exact finite-sample p-values that are robust to small-sample and heteroskedasticity concerns. In our data, RI yields lower p-values for the effect of *Observability* (0.088 versus 0.196 from the asymptotic model), but not for the *Moral message* (0.596 versus 0.670). This pattern is consistent with differences in effect heterogeneity across clusters: when treatment effects are more homogeneous across sessions, the permutation distribution is tighter and the observed estimate appears more extreme, reducing the RI p-value. By contrast, with more variability across sessions, the permutation distribution widens and the RI p-value remains close to the conventional one. In our case, this suggests that reactions to the *Observability* treatment were relatively consistent across clusters, whereas responses to the *Moral message* varied more strongly. We report the full RI results in Appendix A.6, Table 17.[12]

To shed light on perceived descriptive norms, we next examine elicited beliefs about the share of peers who deduct, using the same specification as above.

Pooling all participants, *Observability* increases beliefs by 0.37 pp (p=0.90), while the *Moral message* shifts beliefs downwards by 5.04 pp (p=0.12). Neither of the effects is statistically significant at conventional levels. We then focus on donors, because they are the only participants for whom a deduction decision is observed and thus the relevant group for linking beliefs to behavior. Among donors, *Observability* leads to a change of –9.59 pp (p=0.06) in perceived prevalence. Splitting donors by message clarifies where this comes from. In the *No message* condition, our cleanest read on the effect of visibility and thus active social-norm concerns without any explicit cue, *Observability* lowers beliefs by –19.13 pp (p=0.02), whereas in the *Moral message* condition the estimate is 0.33 pp and indistinguishable from zero (p=0.97) (Figure 8b).

These descriptive belief patterns are transparent evidence consistent in direction with the behavioral estimate reported above for *Observability* on deduction propensity. We do not estimate causal mediation because beliefs were elicited after the deduction decision; any decomposition would be non-causal and is therefore omitted.

Taken together, the evidence from all three experimental settings points to a consistent mechanism: visibility activates social norms that can either suppress or enable the use of tax incentives, while anonymity deactivates them. This cumulative pattern provides the foundation for our broader contribution to behavioral public economics and the literature on the role of social norms in economic decision-making.

---

[12]Using a List et al. (2024)-style cross-fitted AIPW estimator with session-blocked propensities and random-forest outcome models (folds at the session level), we estimate an effect of -11.9 pp (SE 8.41, N=164) among donors. Given the small covariate set (gender, age, degree) and modest N, this machine-learning adjustment is primarily precision-oriented; conclusions are unchanged. For more details, see Appendix A.6, Table 18.

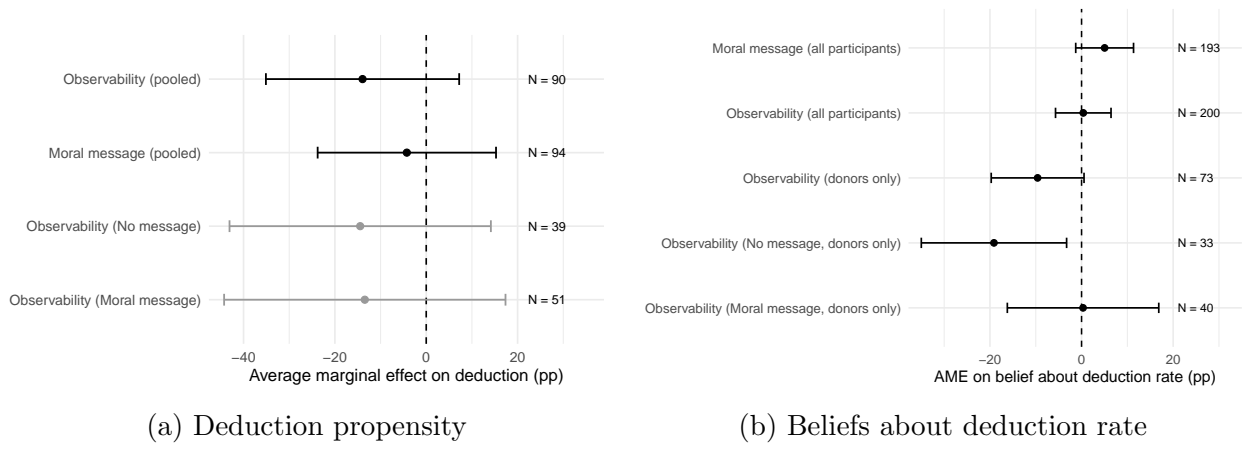(a) Deduction propensity

(b) Beliefs about deduction rate

Figure 8: Treatment effects on behavior and beliefs

Estimates are Lin-adjusted average marginal effects from an OLS linear probability model. The Lin adjustment (Lin 2013) augments the specification with full interactions between each treatment indicator and the covariates (gender, age, education), while retaining the treatment × treatment term (Observability/Anonymity × Moral message/No message) and session fixed effects. Standard errors are cluster-robust at the session level (HC1). Points are average marginal effects (percentage points); whiskers are 95% confidence intervals. Differences in sample sizes result from belief block being added only after the first few sessions.

8   Conclusion

Analysis of Austria's 2017 reform, intended to remove filing frictions, reveals a persistent puzzle. Before the reform, 19.6% of taxpayers claimed a deduction for their charitable donations. Trend-adjusted estimates indicate that the reform raised the deduction share by only 1.8 percentage points, a 9% increase relative to the baseline. At the same time, the average deducted amount fell by about 34.90 EUR, corresponding to a 16.9% decline relative to the pre-reform mean of 206.90 EUR. Even after filing frictions were largely eliminated, the majority of donors still forewent the available tax benefit, and those who did deduct claimed smaller amounts on average. To our knowledge, this is the first systematic evaluation of the reform.

The contrast across donation channels is sharper still. About 86% of donors deduct when giving through the charity's anonymous website, whereas only 0.7% do so in the face-to-face door-to-door drive. A representative survey points to two plausible explanations: limited procedural knowledge and social-image concerns, reflected in a gap between what individuals view as morally appropriate and what they believe others expect. Motivated by these patterns, we designed an anonymous online experiment and two field experiments that directly target information and perceived norms. In the online setting, providing procedural information alone has no effect. Adding a one-sentence norm correction also leaves behavior unchanged, plausibly because deducting is already near saturation, as a baseline rate of 85% leaves little room for further increases at our sample size.

In a socially observable setting (Field Experiment 1), printing the combined information-and-norm paragraph once in local parish journals several weeks before the 2023/24 drive increases the number of deductors in our main specification, active ZIPs, by 0.36 standard deviations, which corresponds to about +7.6 deductors per treated ZIP. Two complementary parish-based specifications and the all-ZIPs variant show smaller, less precise positives. By contrast, an information-only paragraph has no detectable effect across any of the four pre-specified variants.

Field Experiment 2 stress-tests scalability and timing by delivering one-day radio spots during the drive itself in two federal states. Neither the *Morality* treatment in Upper Austria nor the *Information* treatment in Carinthia has a detectable effect on deduction propensity.

Two features likely contribute to these nulls: public-service rules prohibited naming the charity or its event, dampening relevance at the moment of choice, and the campaign was limited to a single day with only a handful of plays, whereas professional guidance suggests sustained multi-week exposure is typically required. A more speculative reading is that norm cues may carry more weight when delivered by local outlets that feel community-proximate, as in parish journals, than by broad state-level broadcasts.

We causally investigate the role of observability in a lab experiment. Participants complete donation choices at their desk and move to a separate terminal for deduction and belief reporting. Under *Anonymity* they are alone at the terminal; under *Observability* a neutral experimenter escorts them and can view the screen. Following our preregistration, pooled donor data show no rejection of the null; the short moral-norm message has no detectable effect either. Because randomization preceded the donation decision, donor composition across the four cells does not balance perfectly. We therefore report average marginal effects from a logit with demographic controls and session fixed effects: *Observability* reduces deduction by 13.94 pp ($p = 0.196$, one-sided), and the moral message remains small and imprecise. The post-hoc minimum detectable effect at 80% power (one-sided, $\alpha = 0.05$) is about 26.8 pp, above the observed estimate. Deduction rates line up naturally with our broader evidence: about 59% under *Anonymity* and 47% under *Observability*, yielding an overall rate of ∼53%, which falls between the online benchmark (∼85%) and the field (∼0.7%).

Taken together, the results point to a common thread about the role of social norms. In settings of full anonymity, norms are effectively deactivated, and deduction rates rise close to their natural ceiling; under such conditions, neither procedural information nor a brief norm cue has room to matter. When deduction is decided in socially observable settings, norms become salient and often suppress take-up. Here, correcting misperceived norms in a credible, community-anchored channel such as parish journals read in advance of face-to-face solicitation increases deduction. By contrast, a one-day, state-wide broadcast does not shift behavior. This null is consistent with the limited number of broadcasting slots, which made exposure easy to miss and left no possibility of delayed treatment; in contrast, a printed journal can be read at any convenient moment before the drive, allowing the message to sink in and to be perceived as a local cue. Across all three studies, the pattern is coherent: what matters is not only whether donors know the procedural steps, but whether social norms are activated, and if so, whether they are corrected in ways that feel credible to the community in which giving takes place.

Our findings on charitable tax-deductibility suggest that the interaction between incentives and social norms is not confined to this setting but generalizes more broadly. Beyond giving, the same mechanisms shape behavior in other domains. Vehicle choice is a salient case: in many countries, large fuel-intensive cars continue to function as markers of status, which may blunt the effect of purchase rebates for electric vehicles. Policy can counter this by reshaping the norm of what ownership signals, for instance through visible public EV fleets or campaigns that portray electric vehicles as modern and aspirational. Home energy investments show a similar pattern. Subsidies for insulation or heat pumps may achieve little because such improvements are invisible to neighbors, whereas rooftop solar spreads once panels become a visible neighborhood standard. Publishing block-level retrofit rates or awarding plaques for energy-efficient homes can therefore complement financial incentives by shifting the local norm. Even the take-up of earned-income credits may be constrained when claiming is seen as stigmatizing. Letters that emphasize how many peers file or testimonials that frame claiming as responsible financial behavior can change the perception of what the action communicates. Across these settings, the broader lesson is consistent: to move behavior at scale, policy must align the private payoff with the social image, either by correcting misperceived norms or by reshaping them directly.

References

Allcott, Hunt (2011). "Social norms and energy conservation". *Journal of Public Economics* 959-10, pp. 1082–1095. 2011.

Andreoni, James, Justin M Rao, and Hannah Trachtman (2017). "Avoiding the ask: A field experiment on altruism, empathy, and charitable giving". *Journal of Political Economy* 1253, pp. 625–653. 2017.

Ariely, Dan, Anat Bracha, and Stephan Meier (2009). "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially". *American Economic Review* 991, pp. 544–555. 2009.

Bénabou, Roland and Jean Tirole (2006). "Incentives and prosocial behavior". *American Economic Review* 965, pp. 1652–1678. 2006.

— (2011). "Identity, morals, and taboos: Beliefs as assets". *The Quarterly Journal of Economics* 1262, pp. 805–855. 2011.

Bhargava, Saurabh and Dayanand Manoli (2015). "Psychological frictions and the incomplete take-up of social benefits: Evidence from an IRS field experiment". *American Economic Review* 10511, pp. 3489–3529. 2015.

Bicchieri, Cristina (2016). *Norms in the wild: How to diagnose, measure, and change social norms.* Oxford University Press.

Bursztyn, Leonardo and Robert Jensen (2017). "Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure". *Annual Review of Economics* 91, pp. 131–153. 2017.

Bursztyn, Leonardo and David Y Yang (2022). "Misperceptions about others". *Annual Review of Economics* 141, pp. 425–452. 2022.

Chen, Daniel L, Martin Schonger, and Chris Wickens (2016). "oTree—An open-source platform for laboratory, online, and field experiments". *Journal of Behavioral and Experimental Finance* 9, pp. 88–97. 2016.

Chetty, Raj (2015). "Behavioral economics and public policy: A pragmatic perspective". *American Economic Review* 1055, pp. 1–33. 2015.

Chetty, Raj, John N Friedman, and Emmanuel Saez (2013). "Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings". *American Economic Review* 1037, pp. 2683–2721. 2013.

Chetty, Raj and Emmanuel Saez (2013). "Teaching the tax code: Earnings responses to an experiment with EITC recipients". *American Economic Journal: Applied Economics* 51, pp. 1–31. 2013.

DellaVigna, Stefano, John A List, and Ulrike Malmendier (2012). "Testing for altruism and social pressure in charitable giving". *The Quarterly Journal of Economics* 1271, pp. 1–56. 2012.

Deutscher Bundestag (2007). *Entwurf eines Gesetzes zur weiteren Stärkung des bürgerschaftlichen Engagements.* Accessed: 2025-04-24. URL: https://dserver.bundestag.de/btd/16/052/1605200.pdf.

Ekström, Mathias (2012). "Do watching eyes affect charitable giving? Evidence from a field experiment". *Experimental Economics* 15, pp. 530–546. 2012.

Exley, Christine (2018). "Incentives for prosocial behavior: The role of reputations". *Management Science* 645, pp. 2460–2471. 2018.

Greiner, Ben (2015). "Subject pool recruitment procedures: organizing experiments with ORSEE". *Journal of the Economic Science Association* 11, pp. 114–125. 2015.

HM Treasury (2015). *Simplifying the Gift Aid Donor Benefit Rules: A Call for Evidence.* Accessed: 2025-04-24. URL: https://assets.publishing.service.gov.uk/media/5a814976%5Callowbreak%20ed915d74e33fd609/Donor_Benefits_call_for_evidence_web_v4.pdf.

Joint Committee on Taxation (2013). *Present Law and Background Relating to the Federal Tax Treatment of Charitable Contributions*. Accessed: 2025-04-24. URL: https://www.jct.gov/getattachment/74f3781f-f57d-4a64-8229-053c53e1d3c8/x-4-13-4506.pdf.

Lane, Tom, Daniele Nosenzo, and Silvia Sonderegger (2023). "Law and norms: Empirical evidence". *American Economic Review* 1135, pp. 1255–1293. 2023.

Lin, Winston (2013). "Agnostic notes on regression adjustments to experimental data: Re-examining Freedman's critique". *The Annals of Applied Statistics*, pp. 295–318. 2013.

List, John A, Ian Muir, and Gregory Sun (2024). "Using machine learning for efficient flexible regression adjustment in economic experiments". *Econometric Reviews* 441, pp. 2–40. 2024.

Republik Österreich (2023). *Government Bill on Tax Deductibility of Donations*. Accessed: 2025-04-24. URL: https://www.parlament.gv.at/dokument/XXVII/I/2319/fname_1596448.pdf.

République Française (2003). *Loi n° 2003-709 du 1er août 2003 relative au mécénat, aux associations et aux fondations*. Accessed: 2025-04-24. URL: https://www.legifrance.gouv.fr/dossierlegislatif/JORFDOLE000017760169/.

Stantcheva, Stefanie (2021). "Understanding tax policy: How do people reason?" *The Quarterly Journal of Economics* 1364, pp. 2309–2369. 2021.

A   Robustness and Heterogeneity

This appendix presents supplementary robustness checks and heterogeneity analyses that support the empirical and experimental findings reported in the main text. Sections A.1 and A.2 revisit the administrative and survey evidence with alternative specifications and additional event–study plots. Sections A.3, A.4, A.5, and A.6 provide further evidence for the online, field, and lab experiments, including subgroup analyses by gender, prior giving, and baseline deduction behavior.

Unless otherwise indicated, we use the same variable definitions and estimation strategies as in the main text. All confidence intervals are two–sided and clustered at the treatment unit (ZIP code or individual) level.

A.1   Empirical Data

This subsection supplements Section 3 with additional detail on online-giving behavior and on the effect of Austria's 2017 tax reform on deduction outcomes.

Tables 6 and 7 present summary statistics by deduction status and demographic characteristics. Figure 9 documents the pronounced seasonal clustering of online donations, justifying our focus on the annual drive period.

Figures 11 and 12 report event–study estimates of the reform's impact on (i) deduction propensity and (ii) average deducted amounts. Tables 8 and 9 present trend-adjusted interrupted time series (ITS) estimates for deduction propensity and the average deducted volume. Further heterogeneity results are provided in Online Appendix B.1.

Identification and trend adjustment. Because the reform was implemented nationally on January 1, 2017, there is no contemporaneous untreated control group. Moreover, the bracket-level event study for the extensive margin (share of deductors) exhibits a clear non-zero pre-trend in 2013–2016, implying that a simple level-shift comparison to 2016 would overstate the reform's impact (Figure 11). To address this, we construct a trend-adjusted counterfactual by estimating, for each bracket cell, a linear trend on the pre-reform years (2013–2016, weighted by the number of taxpayers). We then extrapolate this cell-specific trend to 2017–2019 and define the detrended outcome as the difference between the observed series and the pre-trend prediction. Regressing the detrended series on a post-reform indicator with cell fixed effects (and clustering by cell) yields an average post-2017 deviation from the counterfactual of 1.77 percentage points (s.e. 0.18 pp.) (Table 8). Relative to the pre-reform mean of 19.6%, this corresponds to a 9% increase in the share of taxpayers who deduct donations.

For average deducted amounts, an event study relative to 2016 does not satisfy parallel pre-trends (joint test of 2013–2015 leads: $p < 10^{-26}$). We therefore use a trend-adjusted ITS. For each bracket cell, we estimate a linear trend on 2013–2016 weighted by the number of deductors, extrapolate this trend to 2017–2019, and define the detrended outcome as the difference between the observed series and the pre-trend prediction. Regressing the detrended series on a post-2017 indicator with cell fixed effects and cluster-robust standard errors yields an average post-reform deviation of –34.90 EUR (s.e. 1.70 EUR), relative to a pre-reform mean of 206.90 EUR (Table 9). This corresponds to a –16.9% reduction in average deducted amounts. Figure 13 shows that the downward shift is also visible in the full distribution of average deducted amounts. These ITS estimates correct for the observed pre-trends and attribute the post-2017 changes in deduction outcomes to the reform under the assumption that cell-specific linear pre-trends would have continued absent the policy.
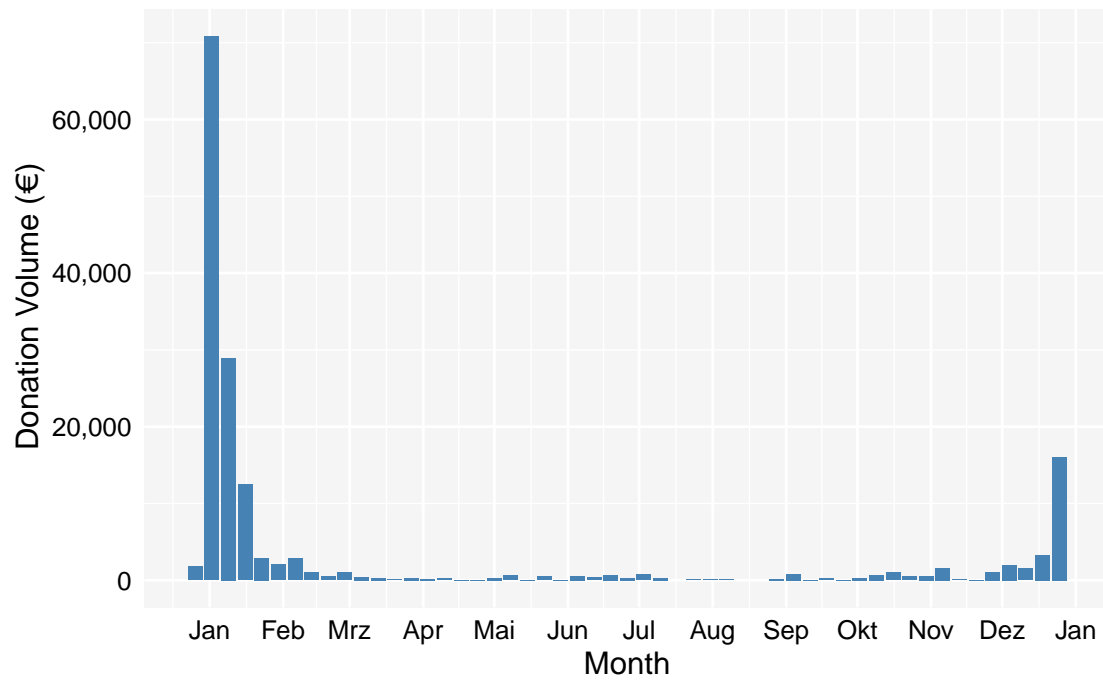
Figure 9: Distribution of incoming donations through the charity's online tool (2022).
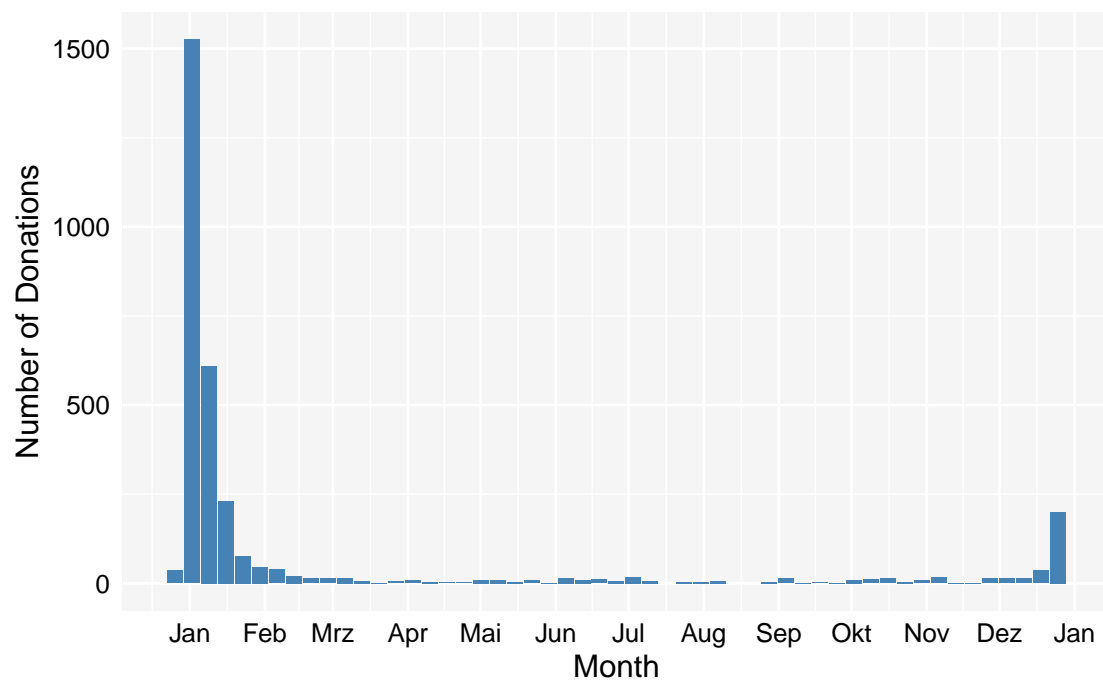

Figure 10: Count of online donations by week (2022).

Deduction status	Observations	Mean (€)	SD (€)
Deductor	16321	54.30	90.00
Non-deductor	2721	36.50	126.10
Full sample	19042	51.80	96.20

[*] Deductor refers to donors who requested to tax-deduct their donation.

Table 6: Summary statistics of online donation data by deduction status.

Gender	N (D)	Mean € (D)	SD € (D)	N (ND)	Mean € (ND)	SD € (ND)
Female	6608	48.65	87.26	1471	32.68	52.59
Male	9086	57.05	84.50	1124	32.49	54.64
Unknown	627	74.12	161.88	126	116.00	528.95

Age group	N (D)	Mean € (D)	SD € (D)	N (ND)	Mean € (ND)	SD € (ND)
≤24	185	41.73	57.15	55	44.08	136.94
25-34	1249	49.73	94.59	154	23.71	20.75
35-44	2812	48.79	80.99	192	29.32	33.47
45-54	3760	54.47	90.18	262	32.48	81.59
55-64	4128	54.34	71.08	270	37.93	100.18
65-74	2667	59.06	105.13	206	38.48	37.13
75+	1243	53.68	74.66	138	37.08	28.47
NA	277	93.78	220.79	1444	38.58	160.25

Federal state	N (D)	Mean € (D)	SD € (D)	N (ND)	Mean € (ND)	SD € (ND)
Burgenland	484	37.90	42.73	57	28.32	27.57
Carinthia	924	44.34	44.76	137	28.49	23.62
Lower Austria	2924	53.64	115.14	310	39.06	97.78
Salzburg	743	50.11	65.64	84	31.80	25.37
Styria	2957	48.89	58.19	435	39.33	239.52
Tyrol & Vorarlberg	1672	55.33	73.43	211	29.10	28.74
Upper Austria	3265	57.79	99.83	262	36.73	59.96
Vienna	3350	61.83	102.69	323	54.26	191.39
NA	2	20.00	0.00	902	31.62	48.74

Note: D = Deductor, ND = Non-deductor.

Table 7: Summary statistics of empirical online donation data.

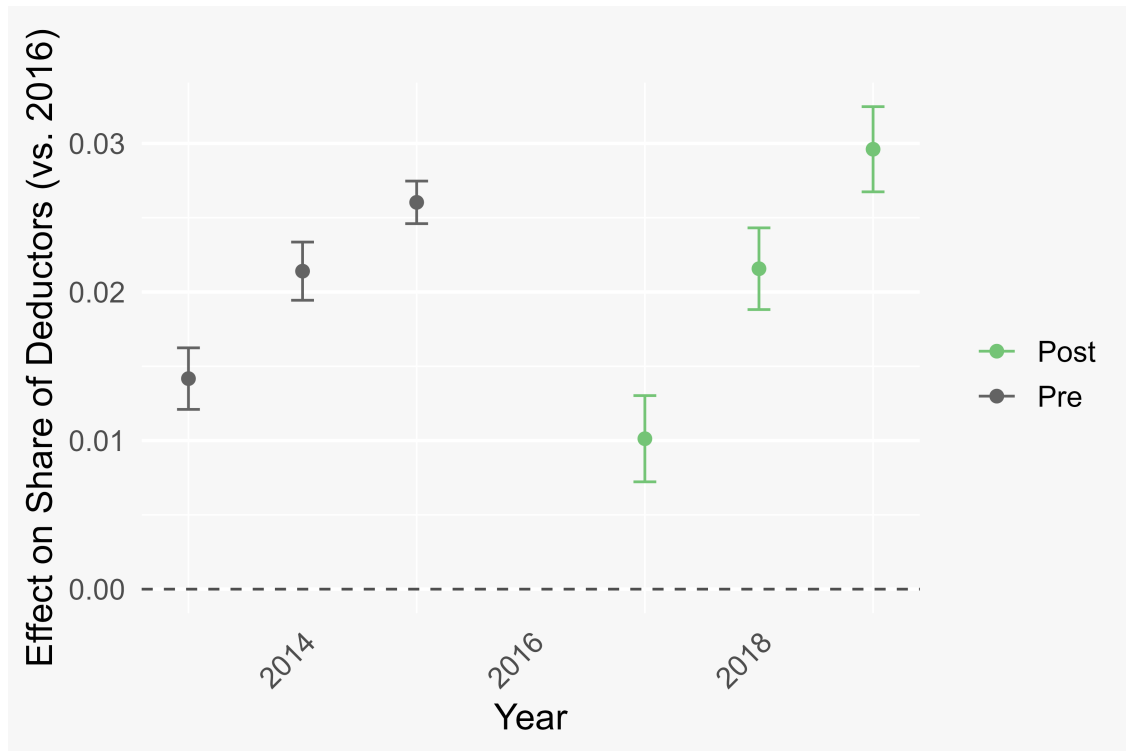Figure 11: Event–study estimates of deduction propensity showing non-zero pre-trend.

Dependent Variable:	Detrended Share of Deductors
Post-2017	0.0177***
	(0.0018)
Observations	68,260
R²	0.52837
Within R²	0.04199
FE	brackets

Clustered (bracket-level) standard-errors in parentheses.
Brackets are defined as low-level aggregates for subgroups.
(e.g.active men of age 20-29 earning 20-25k EUR a year who reside in Vienna.)
* p <0.1, ** p <0.05, *** p <0.01

Table 8: Tax-reform effect on deduction propensity (trend-adjusted ITS).
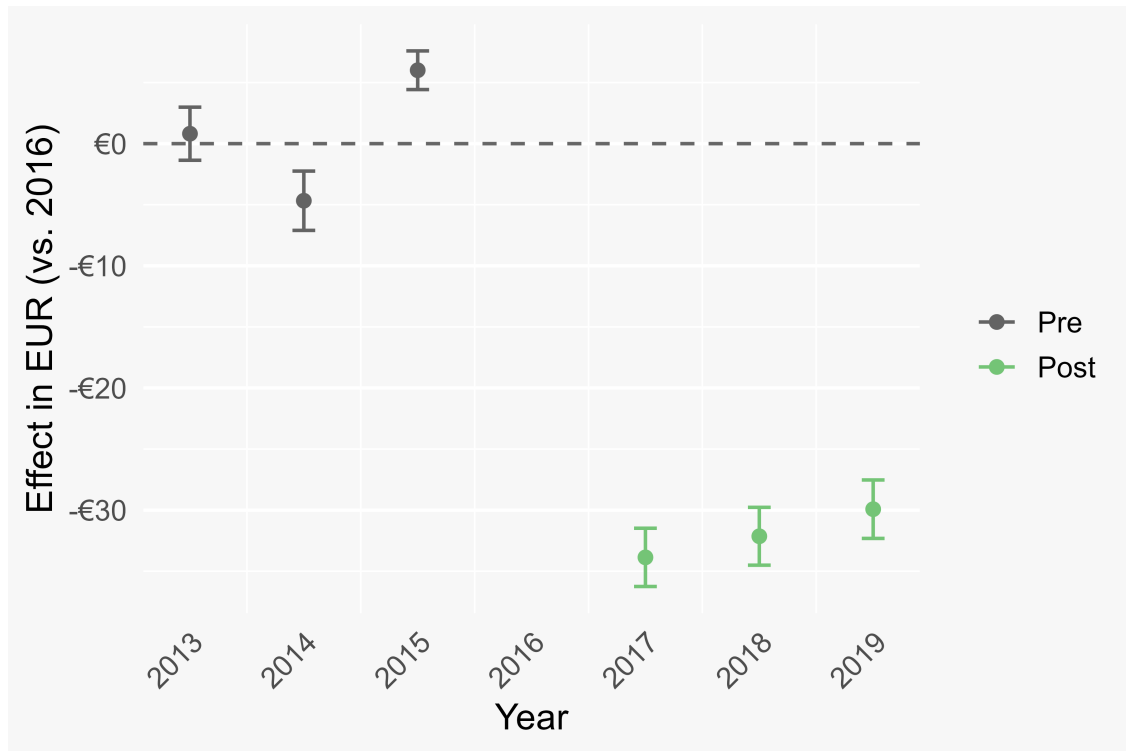
Figure 12: Event–study estimates of average deducted amount showing unstable pre-trend.

Dependent Variable:	Detrended Avg. Deducted Amount (EUR)
Post-2017	-34.90***
	(1.701)
Observations	49,158
R^2	0.44768
Within R^2	0.01667
FE	brackets

Clustered (bracket-level) standard-errors in parentheses.
Brackets are defined as low-level aggregates for subgroups.
(e.g.active men of age 20-29 earning 20-25k EUR a year who reside in Vienna.)
* p <0.1, ** p <0.05, *** p <0.01

Table 9: Tax-reform effect on average deducted amount (trend-adjusted ITS).
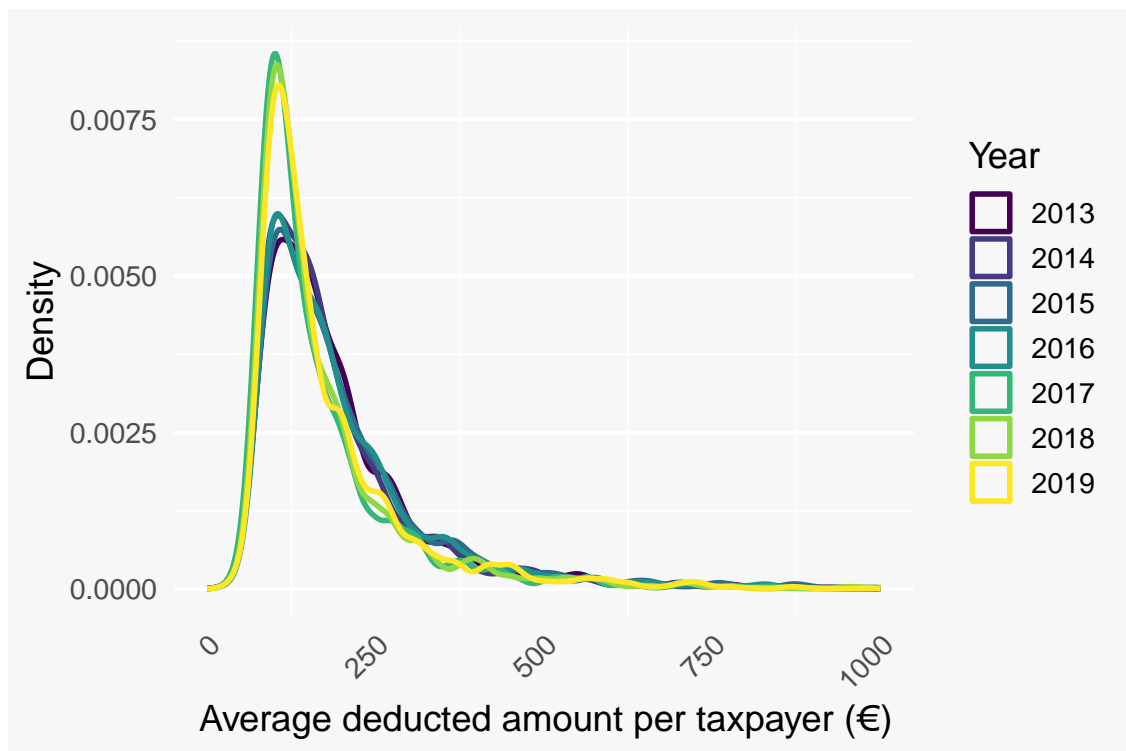
Figure 13: Year-by-year distribution of average deducted amounts, weighted by the number of deductors per cell.

A.2 Representative Survey

This subsection supplements Section 3.3 with additional evidence from the representative survey on taxpayers' beliefs and attitudes toward tax-deductibility of donations.

Figure 14 shows that beliefs about the share of taxpayers who deduct donations differ by respondents' own donation and deduction behavior. Deductors, on average, overestimate the prevalence of deduction relative to non-donors and non-deductors. The actual sample share of deductors was 31.5%.

Figure 15 illustrates flows from first-order moral beliefs to deduction behavior. Respondents who considered tax-deducting donations "appropriate" were substantially more likely to deduct themselves, whereas those who judged it "inappropriate" were predominantly non-donors or non-deductors. This visualizes the close alignment between moral attitudes and actual behavior.

Figure 16 compares the distribution of second-order beliefs (beliefs about others' views) by respondents' own first-order stance. Those who judged deduction inappropriate systematically believed that others would also disapprove, whereas respondents viewing deduction as appropriate held more dispersed second-order beliefs. A Kolmogorov–Smirnov test confirms that the two distributions differ significantly ($p = 0.002$).

Figure 17 documents perceived procedural frictions. Respondents systematically mis-estimated the time needed to claim a deduction, with strong differences across groups (Kruskal–Wallis $p < 0.001$). Deductors reported lower expected time costs than non-donors or non-deductors, consistent with experience and misperceptions shaping beliefs about procedural effort.
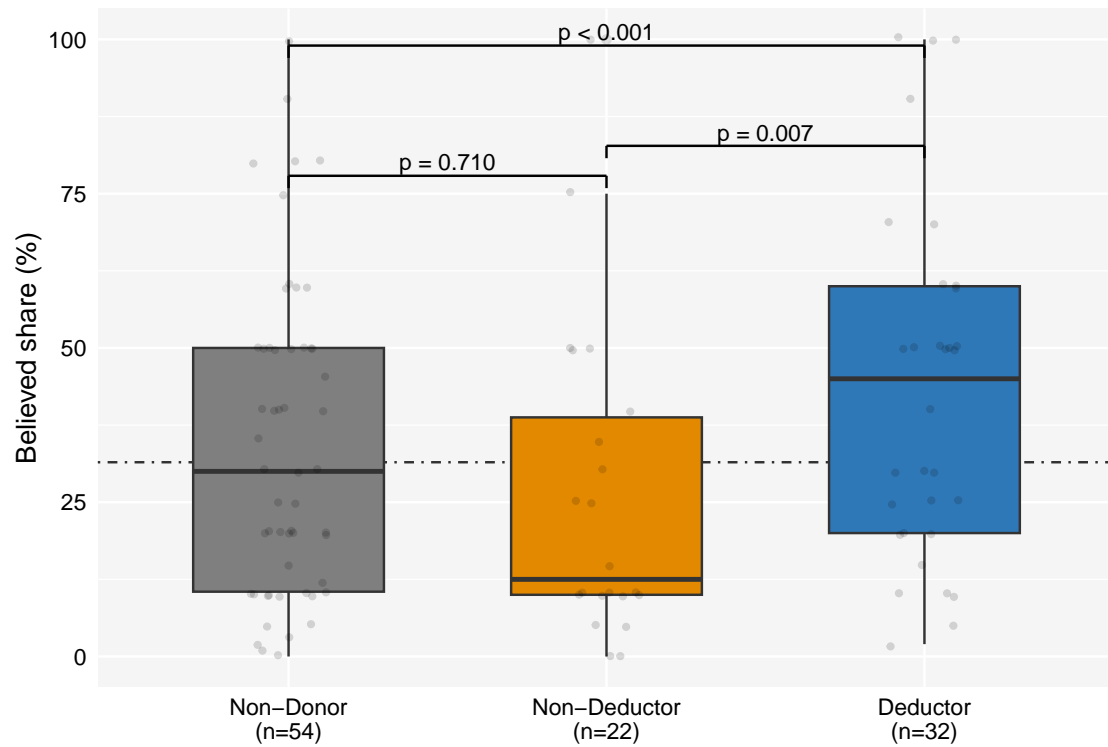
Figure 14: Beliefs about the share of deductors by respondents' own donation and deduction behavior. Reference line marks actual share (31.5%).
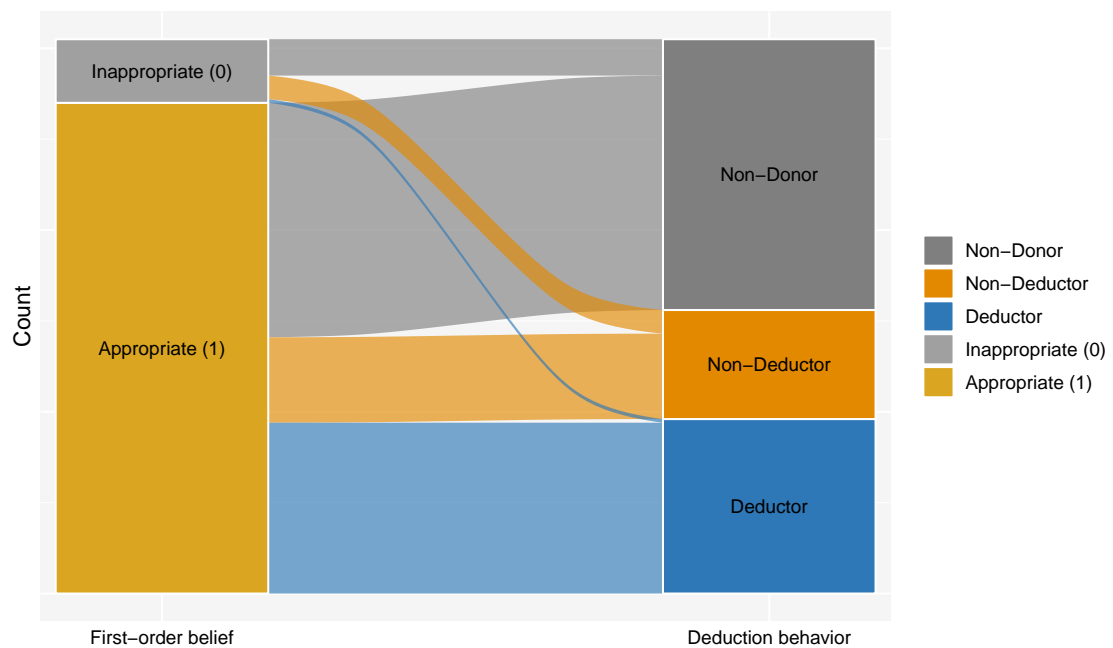

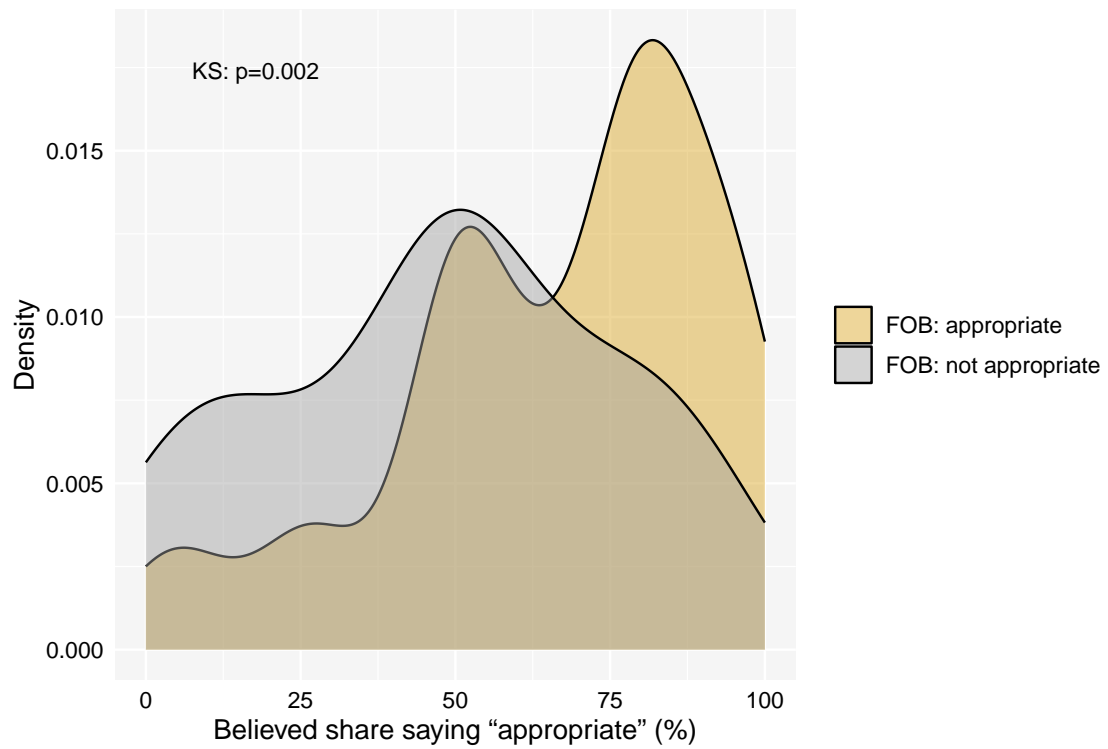Figure 15: Flows from first-order moral beliefs (appropriateness of deduction) to deduction behavior.

Figure 16: Second-order beliefs about others' views by respondents' own first-order stance (Kolmogorov–Smirnov $p = 0.002$).
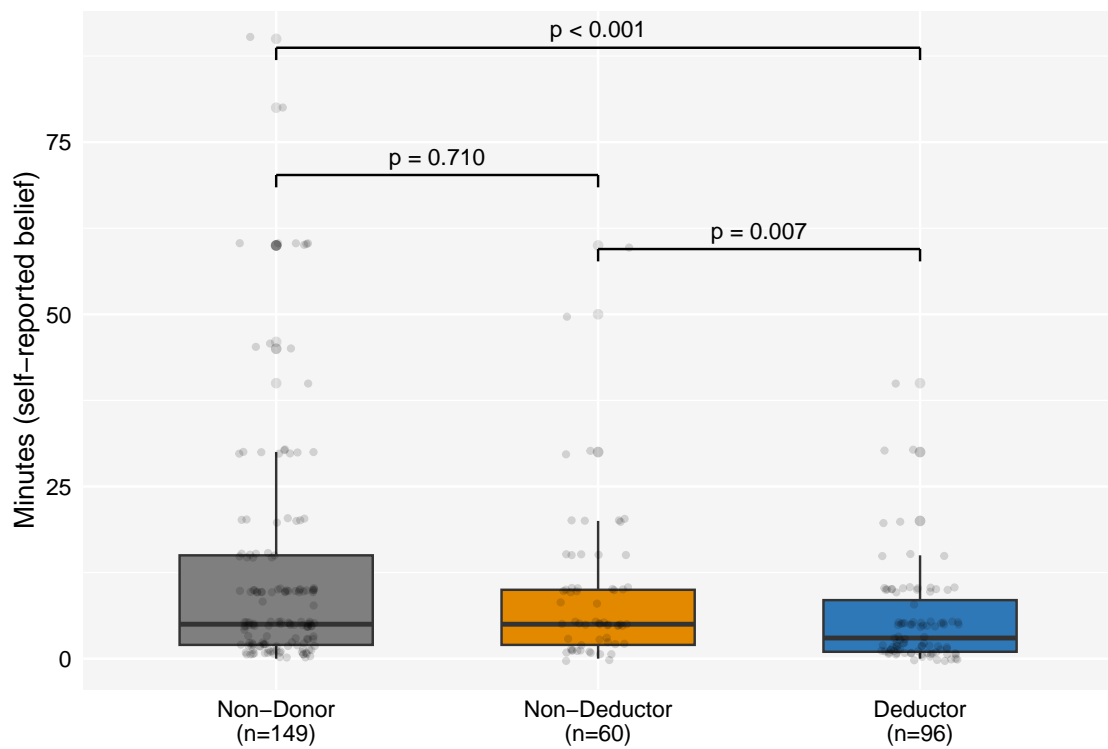

Figure 17: Perceived time required to deduct donations, by respondents' own behavior (Kruskal–Wallis $p < 0.001$).

A.3   Online Experiment

This subsection complements Section 6 with descriptive checks from the online experiment. Figure 18 confirms that treatment messages left donation incidence unchanged—rates are statistically indistinguishable across the *Control*, *Information*, and *Morality* groups. Figure 19 examines allocation across tax-deductible and non-deductible charities. Participants in every treatment donated substantially more to eligible charities, but the share given to eligible causes did not differ by treatment. Finally, Figure 20 shows that among participants who donated to at least one eligible charity, roughly 85–90% claimed the deduction irrespective of treatment, underlining the high baseline take-up rate. Taken together, the evidence indicates that while deductibility shapes allocation choices, our messages did not affect whether or how much participants donated, nor the decision to claim a deduction.

A.3.1   Power Analysis for the Online Experiment

Because only donors to at least one deductible NGO face the deduction decision, the relevant sample sizes are smaller than the full experimental $n = 483$: Control ($n = 94$), Information ($n = 92$), and Morality ($n = 101$). The baseline deduction rate in the Control group was 85%.

We compute the minimum detectable effect size (MDES) for a two-sample test of proportions with two-sided $\alpha = 0.05$ and 80% power. For unequal group sizes n_1 (Control) and n_2 (Treatment), the MDES $\Delta = |p_2 - p_1|$ solves

$$\Delta \;=\; z_{1-\alpha/2}\sqrt{p_{\cdot}(1-p_{\cdot})\left(\tfrac{1}{n_1}+\tfrac{1}{n_2}\right)} \;+\; z_{0.80}\sqrt{\tfrac{p_1(1-p_1)}{n_1}+\tfrac{p_2(1-p_2)}{n_2}},$$

with $p_1 = 0.85$, $p_2 = p_1 + \Delta$, $p_{\cdot} = (p_1 + p_2)/2$, $z_{1-\alpha/2} = 1.96$, and $z_{0.80} = 0.842$.
Results:

- Information vs. Control ($n_2 = 92$ vs. $n_1 = 94$): MDES ≈ 11.8 percentage points.

- Morality vs. Control ($n_2 = 101$ vs. $n_1 = 94$): MDES ≈ 11.6 percentage points.

- Applying a Bonferroni correction for two simultaneous tests ($\alpha = 0.025$) increases the MDES to about 12.5 percentage points.

This means that while the experiment was well-powered to detect large treatment effects, it had limited ability to identify more modest shifts around the already high baseline. The absence of significant differences between treatment and control is therefore consistent with either a true null effect or effects smaller than this detection threshold.
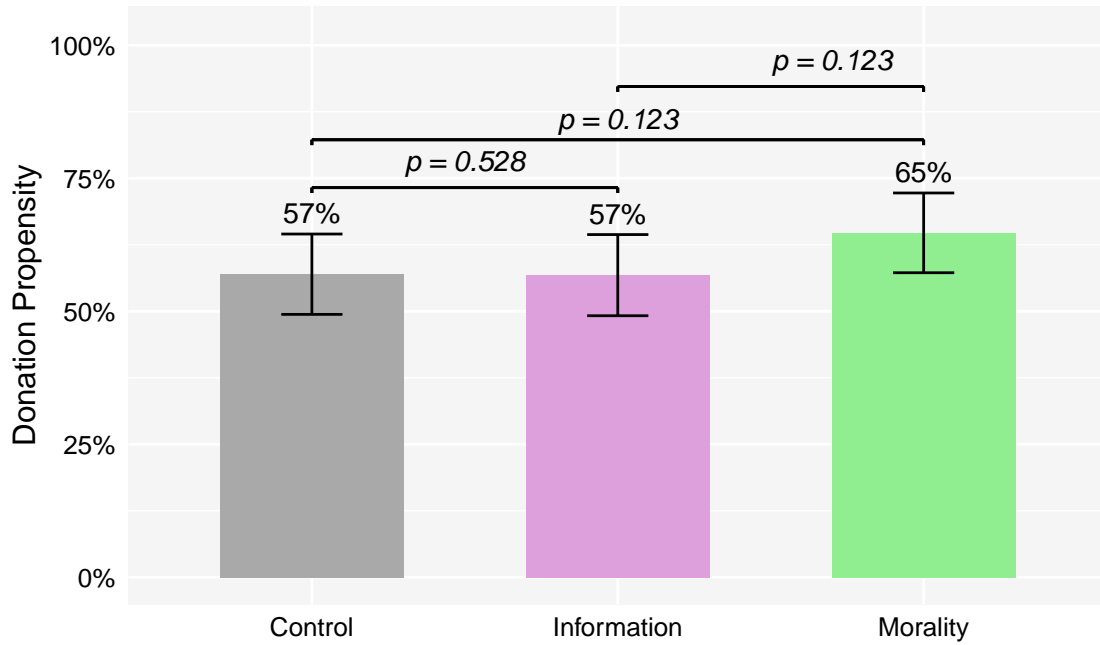
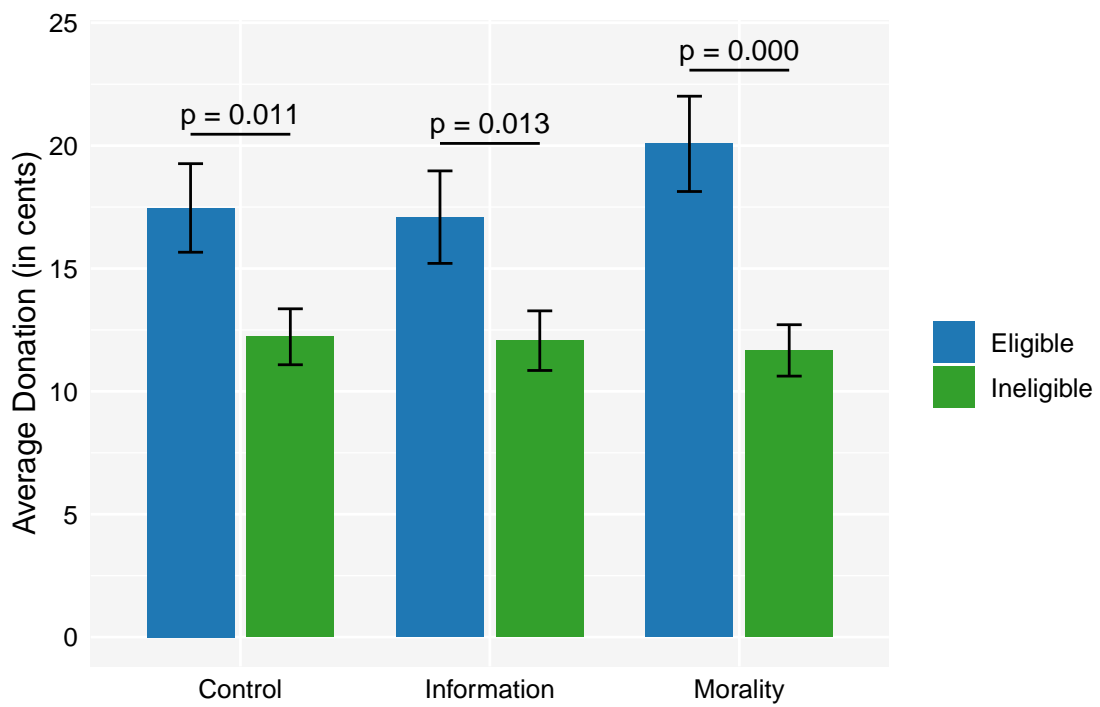Figure 18: Donation propensity by treatment in the online experiment.

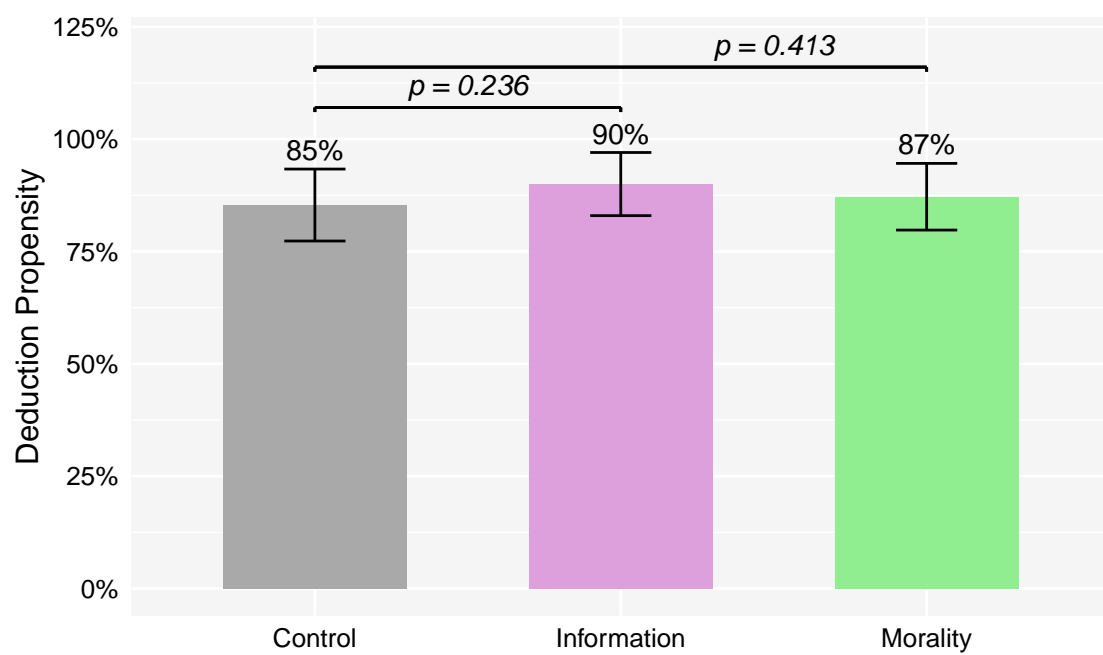Figure 19: Average donation to eligible vs. ineligible charities by treatment.

Figure 20: Deduction propensity among participants who donated to at least one eligible charity, by treatment.

A.4 Field Experiment 1

This subsection expands on Section 4.2 by probing the robustness of the *Information* and *Morality* treatments under alternative aggregation rules. We move from treatment–unit–level analysis (parish-union or parish instances) that covers all Austrian ZIP codes to ZIP-code-level analysis restricted to active ZIP codes—that is, codes with non-zero deduction activity in at least one of the two study years. This switch distinguishes parishes within the same parish union and removes zero-activity cells that dilute treatment effects.

Figures 21 and 22 visualize the distribution of treatment effects in the instance-level data and the adjusted share of deductors for the address-level data.

	ZIP–Year (active ZIPs)	Parish–Year (active)
Information	-4.221	-0.392
	(5.866)	(5.890)
	(0.472)	(0.947)
Num.Obs.	1626	1576
R2	0.828	0.827
Morality	-4.118	-11.774*
	(2.932)	(6.545)
	(0.160)	(0.072)
Num.Obs.	1563	1560
R2	0.826	0.826
Std.Errors	by: ZIP	by: Parish
Fixed effects	ZIP + Drive-Year	Parish + Drive-Year

Only active-ZIP specifications displayed; all-ZIP specifications are identical because ZIPs with no donations have NA volume and are dropped. * p <0.1, ** p <0.05, *** p <0.01

Table 10: Treatment effects on donation volume

The table displays only active-ZIP specifications. All-ZIP specifications are identical because ZIPs with no donations have NA volume and are dropped. * p <0.1, ** p <0.05, *** p <0.01

A.4.1 Minimum Detectable Effect Size at the Address Level

To assess the power of the address-level experiment, we compute the minimum detectable effect size (MDES) under simple binomial assumptions. Each treatment arm contained approximately 2,200 addresses, with a baseline deduction rate of about 0.66%. The standard error for a difference in proportions between two arms is given by

$$SE \approx \sqrt{p(1-p)\left(\frac{1}{n_T} + \frac{1}{n_C}\right)},$$

where p is the baseline probability and n_T, n_C are the number of observations per arm. Plugging in $p = 0.0066$ and $n_T = n_C \approx 2{,}200$ yields $SE \approx 0.00242$ (0.242 percentage
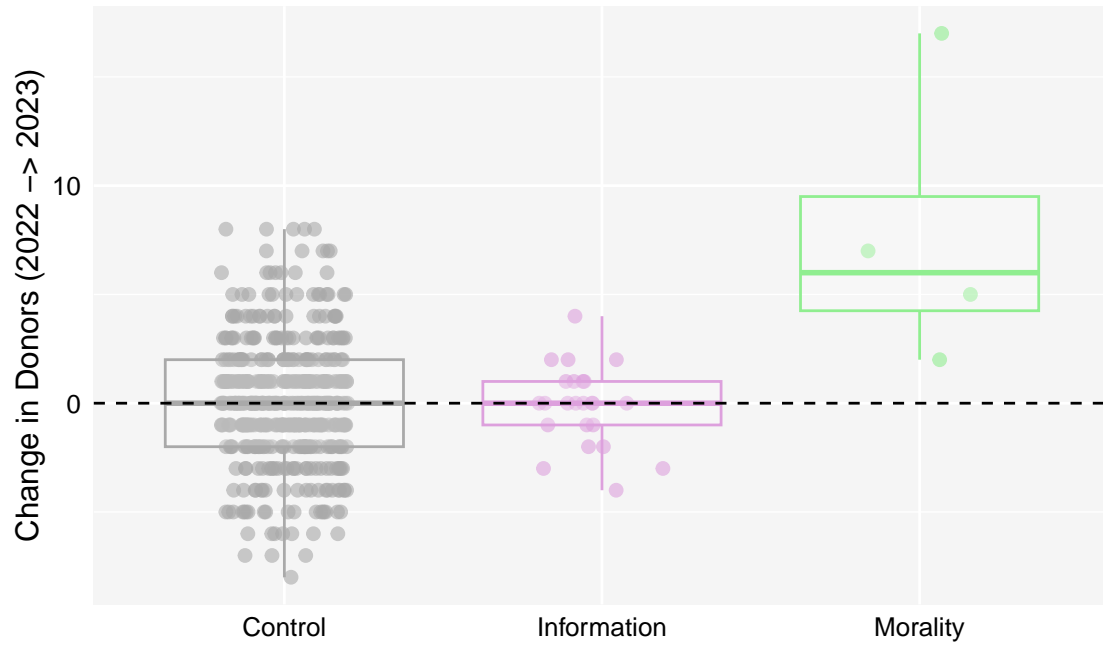
Figure 21: Distribution of treatment effects on the change in deducting donors.
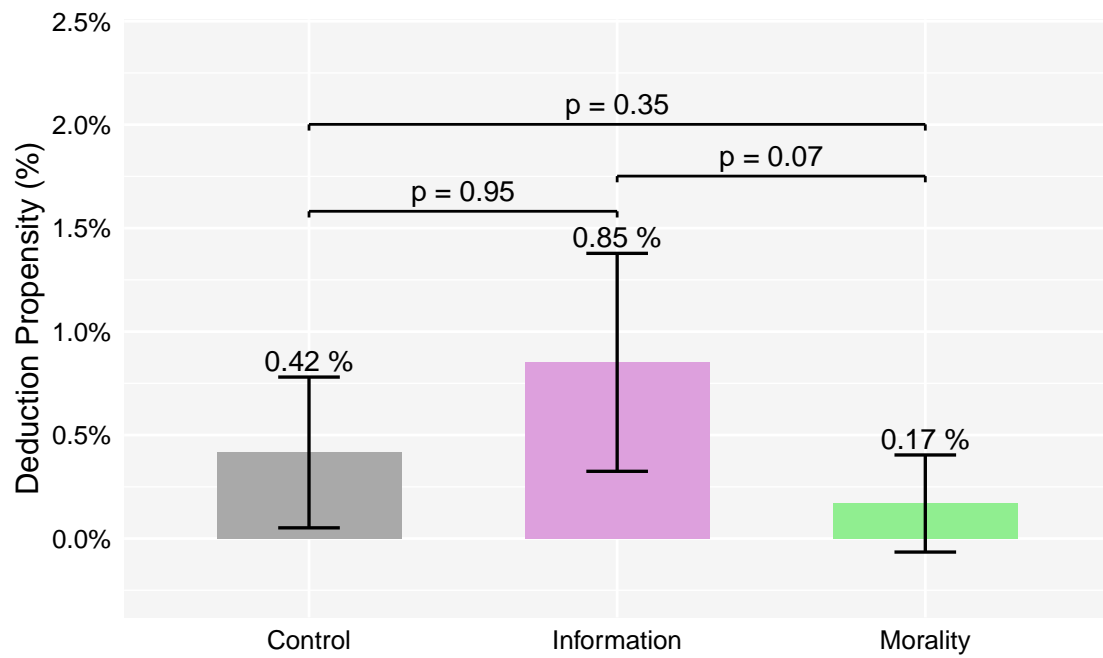

Figure 22: Share of deductors in the address-level data adjusted by donation propensity.

points). A two-sided test at the 5% level with 80% power requires an effect of approximately $(1.96 + 0.84) \times SE \approx 0.0068$, i.e.,0.68 percentage points. This corresponds to about 15 additional deductors in the Morality arm. By comparison, the effect implied by the ZIP-level analysis is only about 0.22 percentage points (≈ 5 deductors). The address-level study was therefore underpowered to detect effects of the magnitude observed at the ZIP level.

A.4.2 *Information* treatment

> For 70 years, children and adults in Austria have been going from door to door as carol singers. Although cars have now become an important means of transport for bridging the sometimes large distances between houses, sleds were still used in the early years depending on snow conditions. The forms of payment have also evolved over time; in some parishes it is now possible to donate by debit card. Yet even though minor aspects of carol singing have continually adapted to the spirit of the times, the core of the campaign has remained unchanged: doing good for people in poor countries around the world. One important change, however, occurred in 2009. Since then, your donations to the carol singers have been tax-deductible. To claim the deduction, simply give your name and date of birth to the adult accompanying the carol-singing group, who then transmits your details directly to the tax office.

The German original can be inferred from Figure 45 in Online Appendix B.4: it is the leaflet text minus *the final paragraph that is unique to the Morality treatment.*

A.4.3 *Morality* treatment

The Morality treatment reproduced the same text as the Information treatment and added the following paragraph:

> Being able to deduct donations even makes it possible to provide greater support to those in need. A recent survey also found that 88 % of respondents consider it morally appropriate to deduct donations.

The full German version, including this additional paragraph, is also visible in Figure 45 in Online Appendix B.4.

A.5 Field Experiment 2

This appendix documents the alternative specifications and robustness checks that underpin the null result in Field Experiment 2 (subsection 5). We report (i) DiD event-time dynamics and alternative DiD estimators, (ii) RDD estimates across bandwidths ($\pm 10, \pm 20, \pm 30$ km) including a 1 km donut exclusion and density/continuity diagnostics, and (iii) placebo tests using pre-treatment periods. Across all exercises, estimates remain small and statistically indistinguishable from zero, reinforcing the main-text conclusion.

A.5.1 Difference-in-Differences (DiD)

Table 4 in the main text (subsection 5) reports the core two-way fixed effects Poisson pseudo-maximum-likelihood DiD estimates. Here, we provide additional specifications and placebo checks:

- Table 11 extends the analysis to the 2022–2025 window.

- Tables 12 and 13 report placebo DiD estimates, documenting the violation of parallel trends in Carinthia in 2024 but not in earlier years.

- Figure 23 visualizes placebo DiD effects across years, and Figure 24 shows dynamic triple-DiD estimates.

	# Deductors (percent change)
Information (Carinthia)	−49.1
	(21.9)
	(0.116)
Morality (Upper Austria)	8.6
	(22.3)
	(0.689)
Num.Obs.	180
Pseudo R2	0.657
Std.Errors	by: date
FE: state	X
FE: date	X

Standard errors and p-values in parentheses. * p <0.1, ** p <0.05, *** p <0.01

Table 11: TWFE PPML DiD — extended sample (2022–2025; state $\times$ date FE; clustered by date).

	# Deductors (percent change)
Information (Carinthia)	−43.6**
	(15.7)
	(0.040)
Morality (Upper Austria)	−0.4
	(13.1)
	(0.975)
Num.Obs.	90
Pseudo R2	0.785
Std.Errors	by: date
FE: state	X
FE: date	X

Standard errors and p-values in parentheses. * p <0.1, **
p <0.05, *** p <0.01

Table 12: Placebo DiD (2024 vs 2023; state × date FE; clustered by date).

	# Deductors (percent change)
Information (Carinthia)	−21.4
	(50.8)
	(0.655)
Morality (Upper Austria)	93.2
	(109.1)
	(0.176)
Num.Obs.	90
Pseudo R2	0.785
Std.Errors	by: date
FE: state	X
FE: date	X

Standard errors and p-values in parentheses. * p <0.1, **
p <0.05, *** p <0.01

Table 13: Placebo DiD (2023 vs 2022; state × date FE; clustered by date).

A.5.2 Regression Discontinuity Design (RDD)

The main RDD results are in section 5.3, Table 5. We complement these with robustness
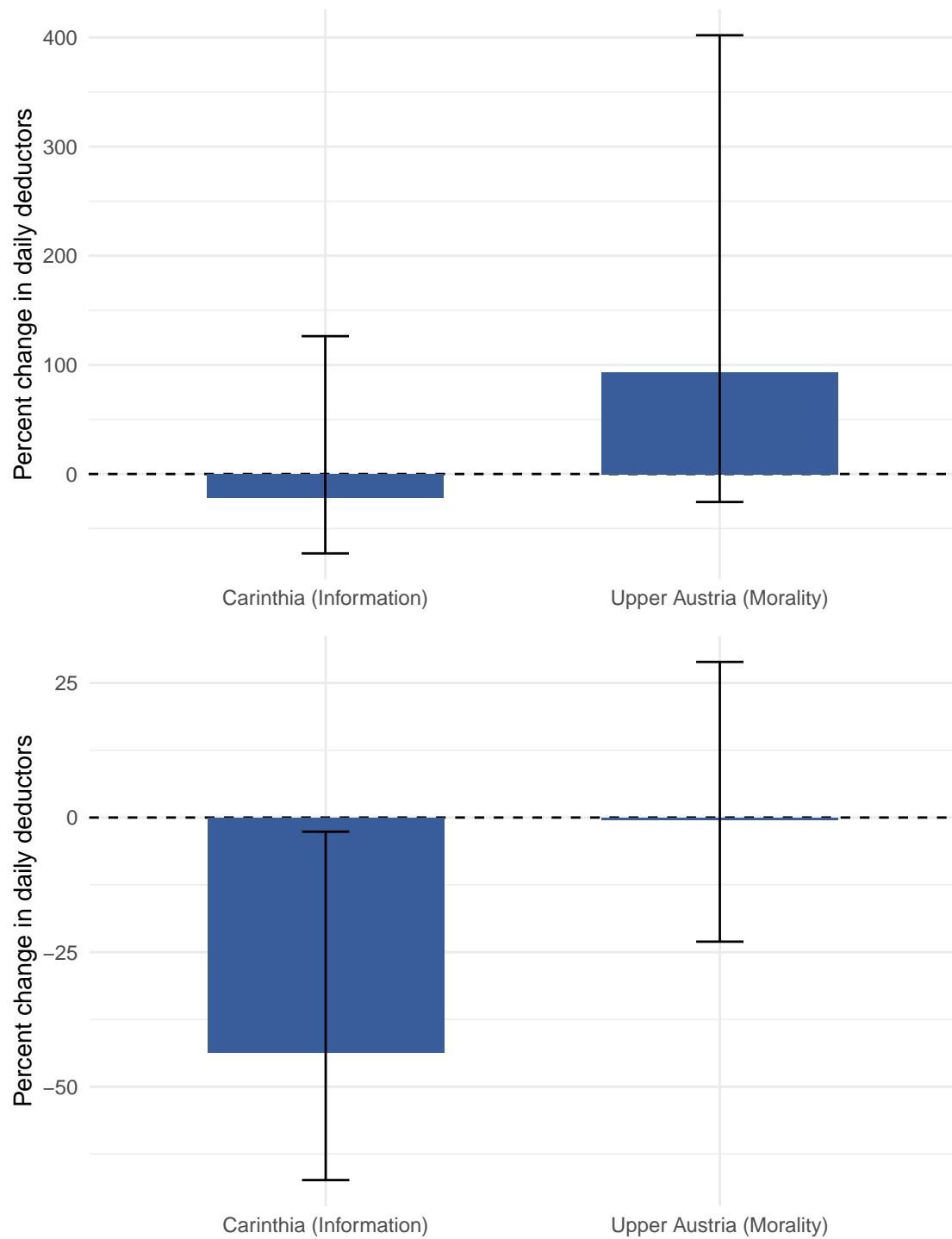checks (donut specification, asinh transform) and alternative bandwidths:

Figure 23: Placebo DiD plots: 2023 vs 2022 (top) and 2024 vs 2023 (bottom).
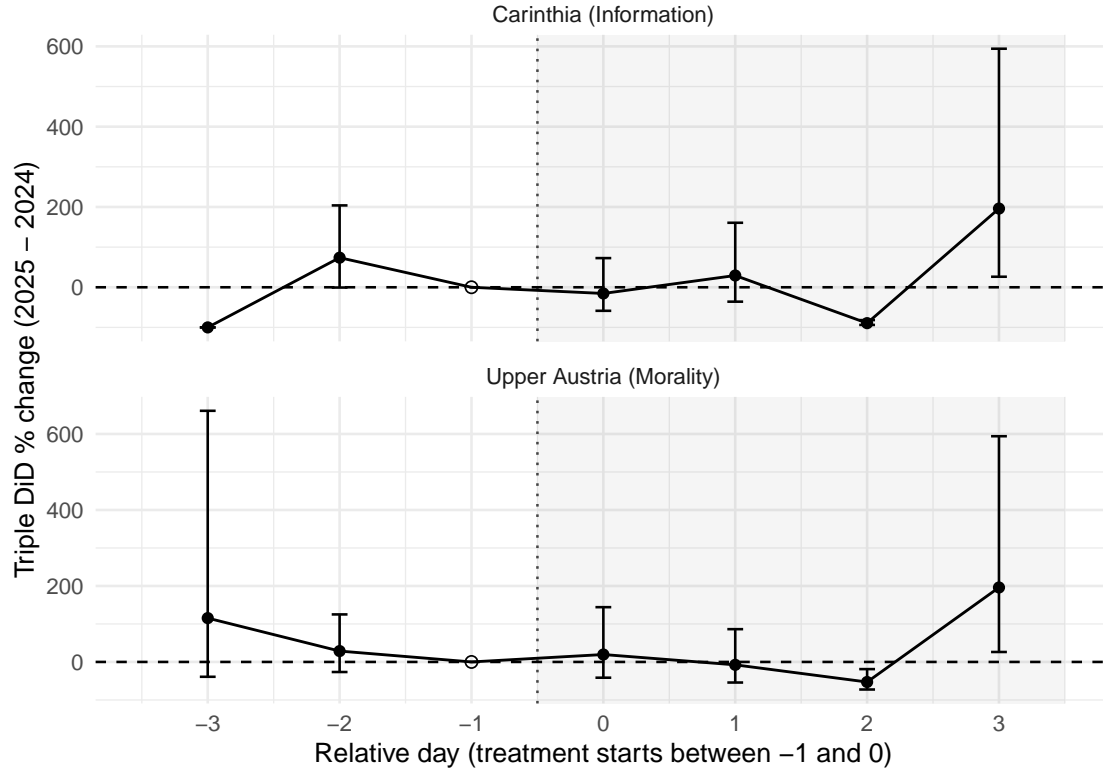
Figure 24: Triple DiD (daily) with state-clustered confidence intervals.

	Information (Carinthia)	*Morality* (Upper Austria)
# Deductors (level)	−6.589*	3.909
	(3.7)	(5.2)
	(0.072)	(0.451)
Num. Obs.	n_L/n_R: 18 / 19	n_L/n_R: 26 / 155

Table 14: RDD robustness checks: donut exclusion and alternative transformations.

Standard errors and p-values in parentheses. n_L and n_R denote the number of observations to the left (outside treatment area) and right (inside treatment area) of the cutoff with 20km bandwidth that are not within a 1km-donut of the broadcast border. * p <0.1, ** p <0.05, *** p <0.01.

	Information (Carinthia)	*Morality* (Upper Austria)
# Deductors (level)	−26.313	3.145
	(16.3)	(3.1))
	(1.000)	(1.000)
Num. Obs.	n_L/n_R: 21 / 22	n_L/n_R: 52 / 181

Table 15: Placebo RDD estimates (2024 only; Jan 2–6).

Standard errors and p-values in parentheses. n_L and n_R denote the number of observations to the left (outside treatment area) and right (inside treatment area) of the cutoff with 20km bandwidth. * p <0.1, ** p <0.05, *** p <0.01.

	h=10km	h=20km	h=30km
Information (Carinthia)	−0.456	−0.485	−0.478*
	(0.5)	(0.3)	(0.3)
	(0.357)	(0.124)	(0.087)
Num. Obs.	n_L/n_R: 7 / 21	n_L/n_R: 18 / 24	n_L/n_R: 31 / 24
Morality (Upper Austria)	−0.030	0.062	−0.001
	(0.2)	(0.3)	(0.2)
	(0.899)	(0.807)	(0.997)
Num. Obs. (n_L/n_R)	n_L/n_R: 16 / 139	n_L/n_R: 29 / 178	n_L/n_R: 40 / 179

Table 16: RDD results across bandwidths (h = 10, 20, 30 km)

Standard errors and p-values in parentheses. n_L and n_R denote the number of observations to the left (outside treatment area) and right (inside treatment area) of the cutoff with indicated bandwidth. * p <0.1, ** p <0.05, *** p <0.01.
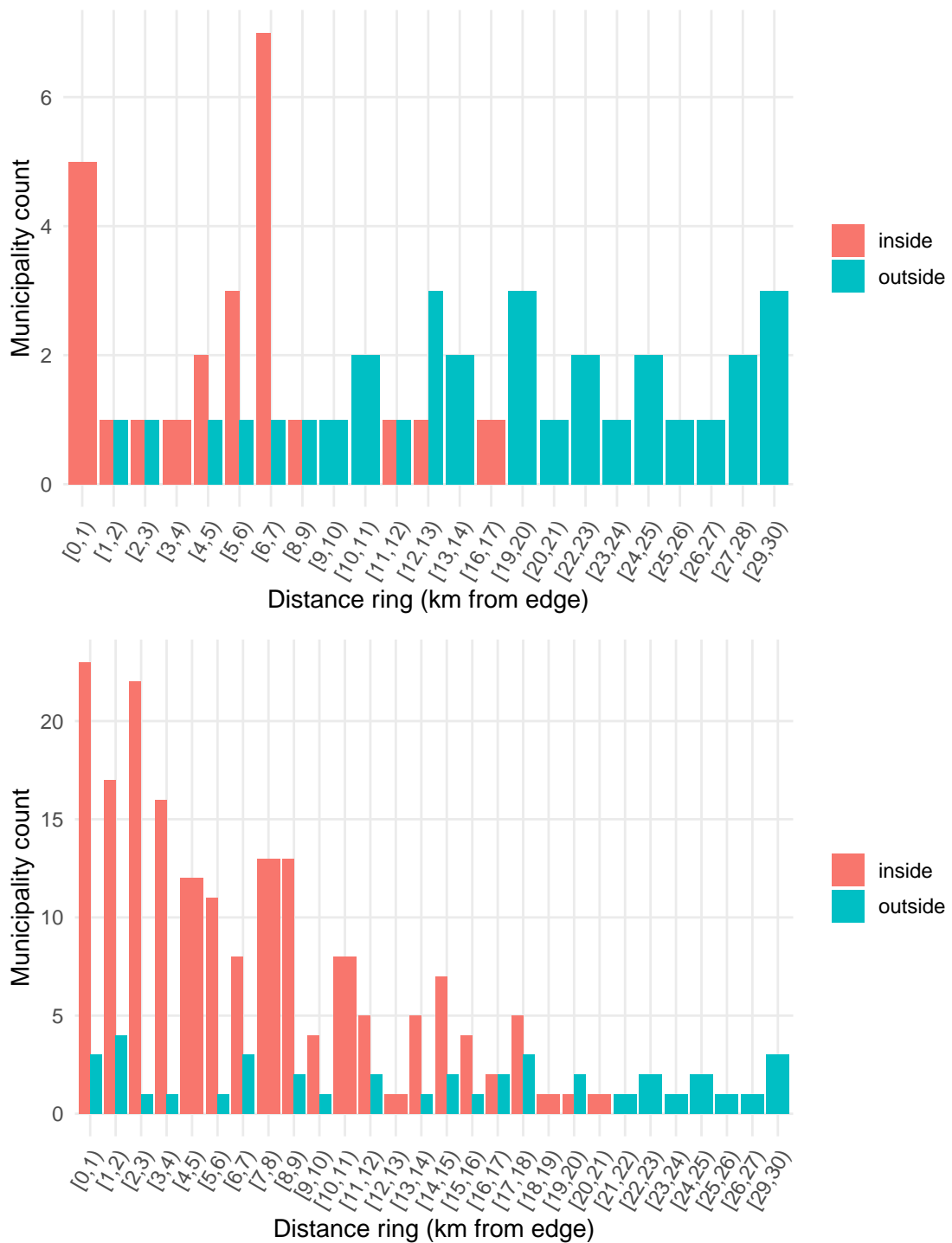
Figure 25: Municipality counts by 1 km ring around treatment cutoff (Carinthia top, Upper Austria bottom).
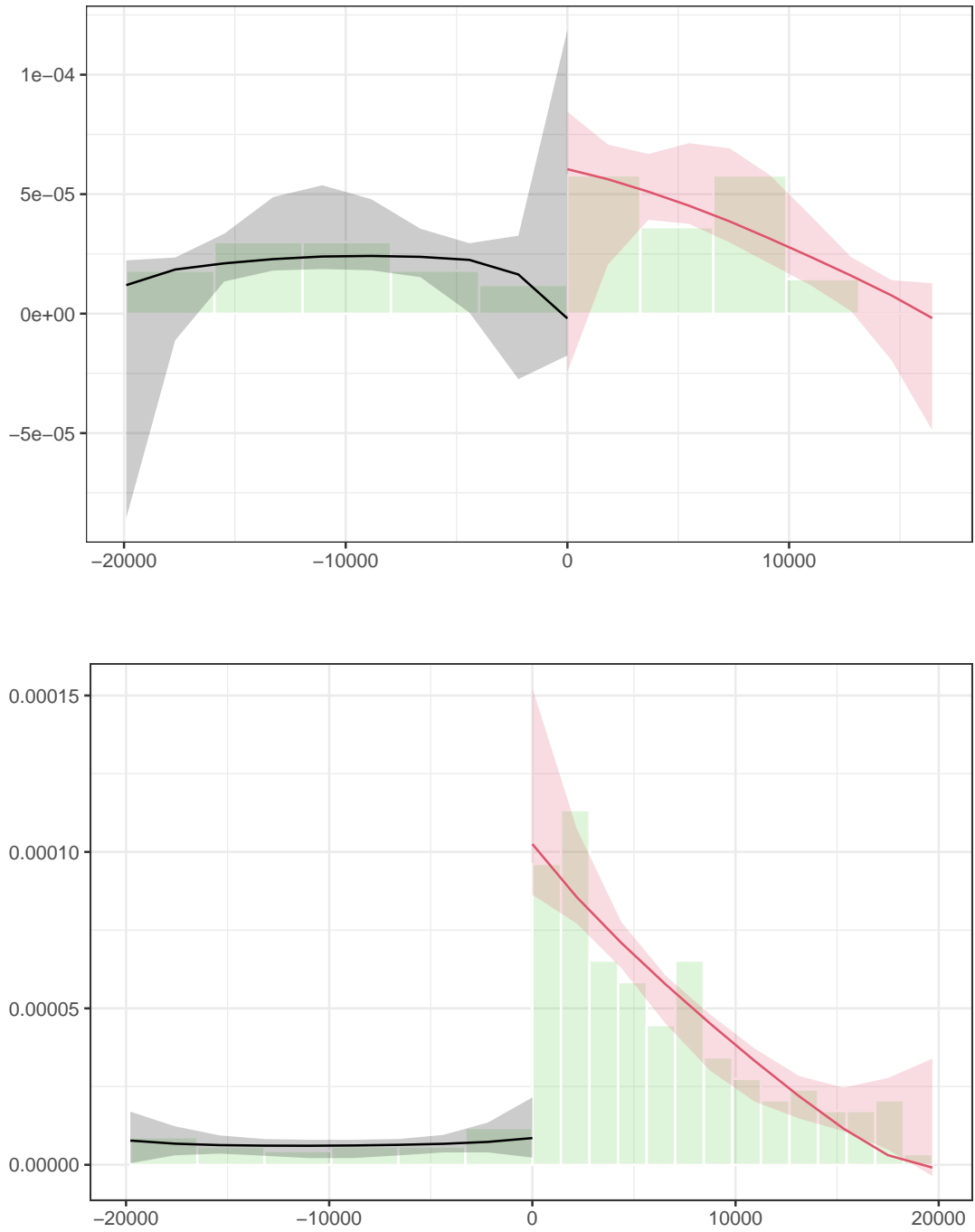
Figure 26: McCrary density tests at the cutoff (Carinthia top, Upper Austria bottom).

We implement McCrary density tests to assess whether observations may have been selectively sorted around the predicted signal cutoff. In Upper Austria, we observe a visible discontinuity, with higher density just above the threshold than below, suggesting potential sorting or measurement error that complicates the validity of the RDD in this state. In Carinthia, by contrast, the density appears smooth across the cutoff, providing no evidence of manipulation. Overall, these results advise caution in interpreting the Upper Austria RDD estimates, while supporting the use of the design in Carinthia.

A.6 Lab experiment

This subsection presents the Lin-adjusted regression results (Table 17) and the List et al. (2024) flexible regression adjustment procedure (Table 18) and supplements Section 7 with balance checks, power calculations, and heterogeneity analyses.

	Deduction propensity — AME (pp)
Observability (overall, Lin-adjusted)	-13.94
	(10.79)
	(0.196)
RI p-value	0.088
Moral message (overall, Lin-adjusted)	-4.25
	(9.96)
	(0.670)
RI p-value	0.596
Num.Obs.	185
Clusters (session)	52
Std.Errors	session
FE	session
Controls	gender, age, degree

Table 17: Lin (2013)-adjusted average marginal effects on deduction propensity

Standard errors and conventional p-values in parentheses. RI p-values are based on 10,000 iterations and appear on the third line below each effect. * p <0.1, ** p <0.05, *** p <0.01

Flexible regression adjustment (AIPW robustness)

As a robustness check we implement a cross-fitted augmented inverse-probability (AIPW) estimator that respects the session-blocked randomization. We split the donors' sample by *session*, fit flexible outcome models $m_w(X) = \mathbb{E}[Y \mid X, W=w]$ on training folds using random forests (covariates: gender, age, degree, and the other randomized arm), predict on held-out folds, and combine these predictions with *session-specific* treatment propensities $p_s = \Pr(W=1 \mid \text{session})$ to form influence scores. Sessions with no within-session variation in treatment are excluded. The ATE equals the mean influence score; we report HC1 standard errors clustered by session. Among donors, this AIPW estimator yields an effect of -11.9 pp (SE 8.41, $N = 164$), consistent in sign and magnitude with the Lin-adjusted estimate reported in the main text. Given the small set of covariates and modest sample size, we view this ML adjustment as a precision-oriented robustness rather than a separate headline result.

	Deduction propensity — AME (pp)
Observability (donors, AIPW-RF, blocked)	−11.91
	(8.41)
	(0.156)
Num.Obs.	164
Sessions	40

Standard errors (clustered by session, HC1) and p-values in parentheses. * p <0.1, ** p <0.05, *** p <0.01

Table 18: List et al. (2024)-based average marginal effects on deduction propensity

Treatment cell	N	Mean age	Share female (%)
Anonymity × Moral message	43	23.8	58.1
Anonymity × No message	52	25.0	59.6
Observability × Moral message	51	24.5	52.9
Observability × No message	39	25.9	51.3

Table 19: Summary statistics of donor characteristics by treatment.

Effect	AME (pp)	SE (pp)	MDE (one-sided, pp)	MDE (two-sided, pp)
Observability	-13.94	10.79	26.8	30.2
Moral message	-4.25	9.96	24.8	27.9

Table 20: Post-hoc minimum detectable effects (80% power, α=0.05; session-clustered HC1).

MDEs assume 80% power and $\alpha = 0.05$. One-sided uses $z_{1-\alpha} = 1.645$, two-sided uses $z_{1-\alpha/2} = 1.96$.
Computed as $(z_. + z_{0.8}) \times$ SE with HC1 cluster-robust SEs (sessions).
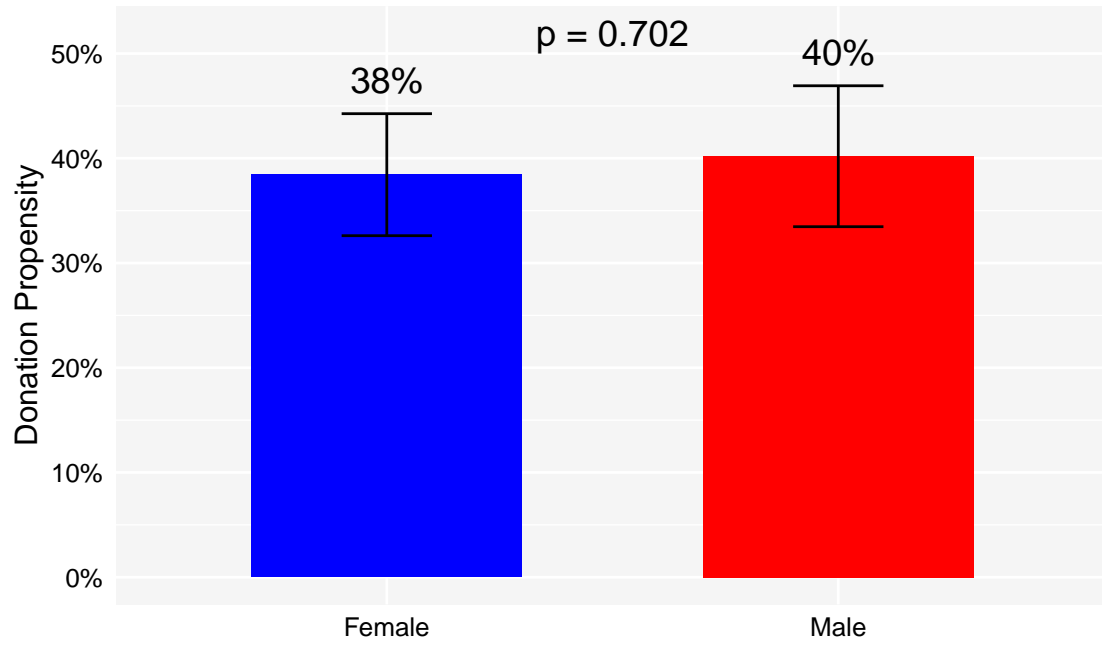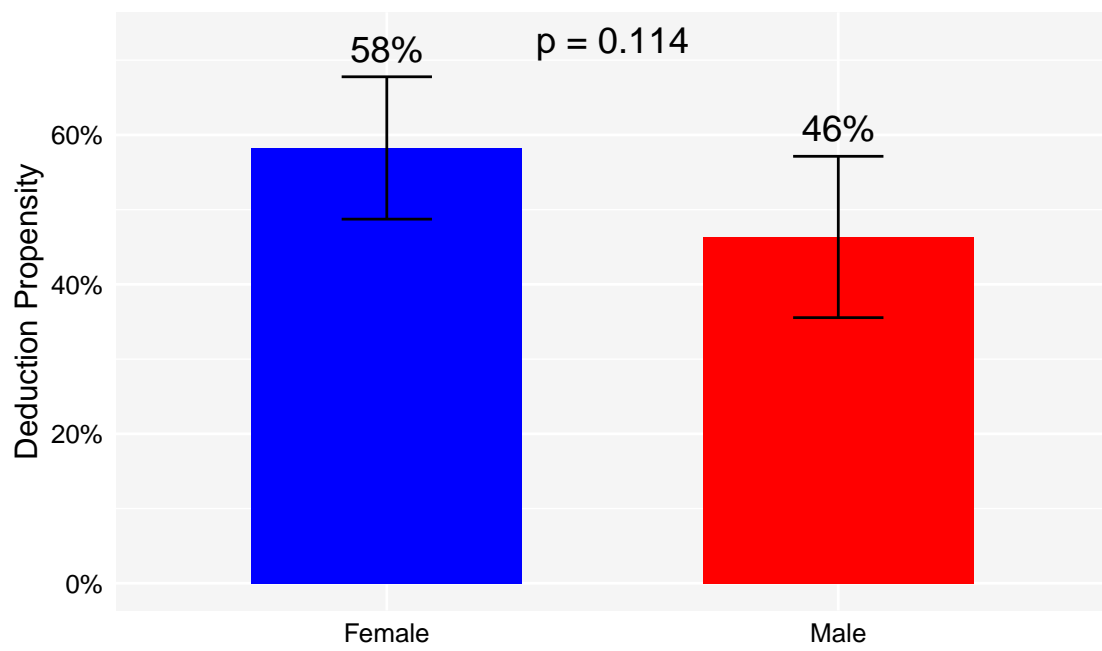
Figure 27: Donation rate by gender across treatments.

Figure 28: Deduction rate by gender across treatments.