Perceptions of (mis)behavior by gender:

Evidence from the Catholic World Youth Day*

Sandra Maximiano† Susana Peralta ‡ João Pereira dos Santos§

António Silva¶

March 2025

Abstract

We provide the first field experiment dealing with prescriptive gender stereotyping on moral issues, during a mass religious event that is likely to make these judgments salient. The survey experiment took place during the 2023 World Youth Day in Lisbon, an event with Pope Francis that gathered 2 million pilgrims from around the World. We asked 1800 Pilgrims from 80 nationalities to judge hypothetical norm-violating scenarios with varying genders, related to lying, sexual promiscuity, and parental care. We find that respondents exhibit high levels of judgment, with women being, in general, less harshly judged than men. We also find that women judge other women less severely in the scenario related to parental care, while men do not discriminate by gender in any scenario. Moreover, more conservative respondents are more judgmental, but do not discriminate between genders, while less fundamentalist ones are less judgemental, but discriminate in favour of women.

JEL Codes: J16, Z12, C93

*The authors are indebted to Miguel A. Fonseca, Diogo Geraldes, Victoire Girard, Muriel Niederle, John de Quidt, and Anna Raute for comments and suggestions. We also thank an army of research assistants for their motivation and outstanding work. João Pereira dos Santos gratefully acknowledges financial support by UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee [grant number EP/Y016718/1]. Pre-registration RCT ID AEARCTR-0011865. Errors are our own.

†ISEG – University of Lisbon. E-mail: smaximiano@iseg.ulisboa.pt.

‡Nova School of Business and Economics. E-mail: peralta@novasbe.pt

§Queen Mary University of London; ISEG – University of Lisbon, REM/ UECE; IZA. E-mail: joao.santos@iseg.ulisboa.pt.

¶ISEG – University of Lisbon. E-mail: apsilva@iseg.ulisboa.pt.

1 Introduction

Society imputes different social and moral responsibilities to men and women, different ways of being and acting, which ultimately shape moral, social, risk, negotiation, and competition preferences (Niederle and Vesterlund, 2011), or perceptions of performance (Bordalo et al., 2019). Moral judgments are shaped by cognitive and emotional processes, which can interact with gender norms and stereotypes (Buckholtz et al., 2015). Glick and Fiske (1996)'s ambivalent sexism theory underscores how stereotypes can affect judgments in scenarios like parental care or sexual behavior, mediated by both hostile and benevolent sexism.

Religions play a significant role in shaping these frameworks, since they provide explicit moral guidelines to tell right from wrong behavior, frequently emphasizing distinct gender roles for sexual behavior, parental care, and honesty (Gundersen and Beck, 2020). Moreover, religion is a powerful socializing force that reinforces group norms and in-group solidarity, which can amplify biases against individuals perceived as deviating from prescribed norms.[1] In the context of this study, Catholicism emphasizes traditional family structures, sexual morality, and parental responsibilities, providing a fertile ground for exploring how religiously informed moral schemes intersect with gender discrimination. By focusing on a highly religious sample, this research uniquely captures the influence of religion on gendered moral judgments in a real-world context.

Gender stereotypes can be both descriptive – beliefs about men and women's characteristics, i.e., how the are or how they behave – or prescriptive – beliefs about how men and women should behave or be. Importantly, prescriptive stereotypes often imply penalties for norm violation. Note that these need not be negative; e.g. stereotypically perceive women as more likely to be caring is a positive bias. These concepts have been widely studied in the psychology literature. Several papers analysing labour and leadership relationships apply these concepts, which lend them self naturally to these settings, as the male-prescribed char-

[1] Becker et al. (2021) review the literature on religion in economic history and discuss the role of human capital. In addition, Sinding Bentzen (2019) discusses the importance and persistance of religion in coping with natural disasters.

acteristic of *agency* is associated with leadership roles. Applications include Heilman et al. (2004, 2024) or Rudman and Glick (2001), which show that women are penalized when they fail to comply with the norms. These penalties may include lower wage recommendations (Heilman et al., 2004), or restricted access to social networks (Casciaro and Lobo, 2005).

We provide the first field experiment dealing with prescriptive gender stereotyping on moral issues. Moreover, we do so in a context of a mass religious event, in which prescriptive stereotyping is prominent. We examine gender discrimination in moral judgment through a field survey experiment conducted during the 2023 World Youth Day in Lisbon, an event that gathered an estimated 2 million pilgrims, for one week, to see Pope Francis in August 2023. Using a diverse and religiously motivated sample of 1800 Catholic pilgrims of 80 different nationalities, we explore how gender shapes moral evaluations of hypothetical scenarios involving lying in the job place, sexual promiscuity, and parental care. Our treatment consists in varying the gender of the individual who pursued the described behaviour, in otherwise standardised fictional scenarios. We are also interested in the impact of the gender of the person who judges the scenario. This context allows us to capture moral judgments in a highly value-oriented and culturally specific setting, offering insights that extend beyond controlled laboratory experiments.

Our main results are as follows. First, respondents exhibit high levels of judgment. Second, women are, on average, less judged than men. Third, while men do not discriminate according to the gender of the person pursuing the action in the hypothetical scenario, women tend to be more lenient about other women's behaviour. These effects are driven by the parental care scenario, a finding which is consistent with the conclusions of the literature about gender discrimination in the labour market (Bertrand and Duflo, 2017) and to the child penalties related to gender and parenthood (Kleven et al., 2019, 2024). Moreover, more conservative respondents are more judgmental, but do not discriminate between genders, while less fundamentalist ones are less judgemental, but discriminate in favour of women. We also present some evidence that the gender of the enumerator does not play a role,

suggesting that experimenter demand effects are not very relevant in this setting (Zizzo, 2010; De Quidt et al., 2018, 2019).

While there is a breadth of theoretical work on the gendered dimensions of moral judgment – encompassing academic, historical, and religious texts – there is scarce empirical work exploring how people judge moral situations differently based on the gender. Related research shows that gender stereotypes affect the way people approach, interpret, and remember information about themselves and others (Ellemers, 2018).

Our paper is related to the nascent experimental literature on prescriptive stereotyping (Bordalo et al., 2016) and misperceptions about others (Bursztyn and Yang, 2022). Bursztyn et al. (2020) show that (upward and truthful) correction of mens' beliefs about their male peers' acceptance of women working outside the home increases their willingness to support their wives seeking outside-the-home jobs. Conversely, women who learn about the true acceptance amongst males are more likely to take higher-paying outside-the-home jobs. Indeed, misperceptions about the true distribution of gender norms are common and context-specific (Bursztyn et al., 2023). More precisely, support for equality-promoting policies amongst men is underestimated in less equal countries. Dhar et al. (2022) analyses a randomized classroom experiment entailing discussions about gender equality, and finds a significant and long-lasting decrease in support for regressive attitudes.

Contrary to these two papers, our treatment does not aim at curbing the stereotype with information provision; rather, we focus on the impact of the gender of the person who judges females or males who fail to comply with prescriptive stereotypes usually associated with women. This approach resembles that of Asmat et al. (2024), who assess the extent to which females are harmed when being judged by females using data from 133 years of piano competitions; however, their analysis is about performance and is unrelated to stereotyping. Moreover, they do not vary exogenously the gender of the judged individual, as we do in our field experiment.

There is also a small, but important, literature about the impacts of mass gatherings,

or other religion-related special events, to which our paper speaks. Bassi and Rasul (2017) documents the influence of the papal visit to Brazil in 1991 on short-run intent to contrapet and long-term outcomes related to fertility. Exposure to papal speeches, which drew heavily on fertility issues, decreased the prevalence of contraception and increased births by 1.6 percent 9 months later. Farina and Pathania (2020) shows that papal visits to Italian provinces decrease abortions by between 10 and 20%, in the period between three and 14 months after the visit. Clingingsmith et al. (2009) exploit the random assignment of a Visa lottery in Pakistan to participate in the annual Haij Muslim pilgrimage to Meca. They find significantly positive treatment effects on global Islamic religious practices, on inter-group acceptance, and on ecumenic beliefs beliefs about peace and equality. They also find a positive effect on acceptance of female education and employment.

The policy relevance of this topic lies in its implications for fairness and equality. Moral judgments are not made in a vacuum; they often inform decisions in governance, law, education, and healthcare. For example, biases in moral judgment could affect jury decisions, hiring practices, or parental custody rulings. As these moral judgments are likely to affect individual preferences, they also can have an impact on education and professional choices, for example, leading men and women to choose different fields of study and careers. Evidence of systematic gender discrimination in moral evaluations calls for nuanced interventions that mitigate these biases, contributing to the development of fairer institutional frameworks.

This paper is organized as follows. In Section 2 we discuss the institutional setting and describe the data. In Section 3 we explain the experiment while in Section 4 we present the results. Section 5 concludes.

2 Institutional Setting and Data

World Youth Day (WYD) is a global event organized by the Catholic Church that gathers young people from around the world, established by Pope John Paul II in 1985. It is held

every two to three years in different cities and features a variety of spiritual, educational, and cultural activities, including masses, catechesis sessions, and a final vigil and Mass with the Pope. The 2023 WYD was held in Lisbon. For this experiment, we took advantage of a massive installation of 150 private confession cabins in a public park in Lisbon – the so-called *Parque do Perdão* (Reconciliation Park), where we approached pilgrims.[2]

This place was especially suited to conduct our survey experiment as this was a space of healing, gift, and joy, where social and moral preferences can be more salient and, importantly, where a transmitting effect of generosity for the general population may arise. In *Parque do Perdão* there were several waiting moments, given the demand and the number of existing concessionaires, contributing to a very generous willingness from the pilgrims to participate in our study. In other spaces, where other social and religious activities took place, our interviews could imply the interruption of the pilgrims' attention, and we did our best to be as non-invasive as possible. Furthermore, it should be highlighted that the assistants who carried out the interviews were trained by the research team so that they are careful and sensitive in their approach, so as not to disturb pilgrims during their most introspective moments.

We restricted participation to individuals with at least 18 years of age, making sure that they understood and spoke Portuguese and/or English.[3] Subjects were incentivized with 22 lotteries (two 500 euros and twenty 50 euros) to increase attentiveness that were later randomly assigned by the research team and announced on the X-Lab website.[4]

We implemented a one-sheet, two-page survey, in A4 format, that collected information on a battery of sociodemographic questions about the respondent, such as age, gender, nationality, education level. Moreover, we ask questions about religious behaviours, such as church going, or praying, that allow us to use the centrality of religiosity scale (Huber and

[2]Confessions were available in the 4 official languages of the WYD Lisbon 2023 (Portuguese, English, Spanish and Italian) and in other languages. More information about this park can be found in `https://www.lisboa2023.org/en/forgiveness-park`.
[3]All participants signed an informed consent agreement stating that they could give up at any point in time.
[4]More information in `http://xlabonline.com/field-experiment-at-world-youth-event-2023/`.

Huber, 2012). We complement these with respondents' views on topical issues in current religious debates, such as the possibility that females preside over the Mass or the extension of religious marriage to same-sex couples. The treatment was implemented on the same paper-based survey. We turn to it below.

We present the descriptive statistics in Table 1. Our sample collected questions from more female respondents (58.2%) than male. Although there were more women research assistants (17) than men (15), male assistants collected slightly more than half of the valid survey responses that we use in our sample. Moreover, our sample is young (on average, 27 years old), highly educated (66% self-reported to have a bachelor degree), was in Lisbon specifically for the WYD and came from abroad. The number of survey responses was evenly distributed on the four days that *Parque do Perdão* operated, with more responses recorded on August 2, when the park remained open later than on the other days.

Table 1: Summary statistics

	Mean	Std. dev.	Min	Max
female respondent	0.582	0.493	0	1
female assistant	0.473	0.499	0	1
age	27.286	10.548	18	79
in Lisbon for WYD	0.890	0.313	0	1
religiosity index	4.541	0.527	1.6	5
foreigner	0.713	0.452	0	1
graduate	0.658	0.474	0	1
August 1	0.230	0.421	0	1
August 2	0.274	0.446	0	1
August 3	0.242	0.429	0	1
August 4	0.253	0.435	0	1

Our sample comprises respondents from 80 nationalities, as shown in Table A1 in the Appendix. The top 5 nationalities are Portuguese (29%), Spanish (12%), Brazilian (8%), French (8%), and American (6%).

3 Treatment

We aim at measuring the gendered dimension of moral judgments of prescriptive stereotypes. To do so, we create three social norm-violating hypothetical scenarios and vary the

gender of the individual taking the actions in each scenario. We then ask the respondents to morally rate the scenarios.

Our (pre-specified) outcome variable is the sum of the responses to three questions, each related to one of the scenarios. In what follows, this variable is called *index*. More specifically, we ask participants to "Morally rate the following fictional situations on a scale from 1 (morally objectionable) to 6 (not morally objectionable):"

1. Alice/ Henry missed work because she/ he went to a party the day before, and lied to the employer saying she/ he was sick

2. Diana/ Adrian always works late and can rarely pick up the children from school

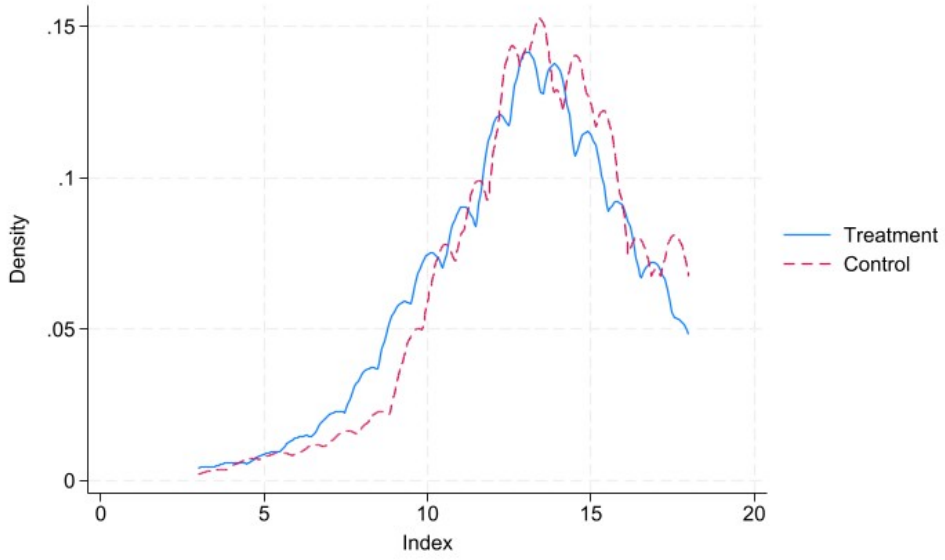
3. Caroline/ Carl, who is married, went out at night and had sex with a stranger

The respondents saw only three female or three male names. Therefore, this is between-subject design as each participant is only exposed to one treatment environment. As such, the participants are not provided with clues to infer the treatment they are assigned to (Charness et al., 2012). Moreover, we are exploiting implicit discrimination (de Quidt, 2024). The treatment refers to female characters undertaking the actions under each of the scenarios.

We prepared a pile of surveys alternating male and female characters in the three hypotethical scenarios. Each of the 32 enumerators in the team took surveys from this common pile and scattered randomly around the Reconciliation Park to recruit subjects. The fact that all the enumerators drew the surveys from the common pile ensures that, on average, one half of the subjects were treated with hypothetical scenarios involving female characters. We show that treatment is balanced on several variables in Table A2 in the Appendix.

The non-parametric distribution of the index is displayed in Figure 1, suggesting first order stochastic dominance of the distribution of the control (i.e., male characters in the hypothetical scenarios) observations over treatment (i.e., female characters in the hypothetical scenarios) ones.

Figure 1: Distribution of responses, by treatment



4 Experimental results

We now present our main results. We begin by estimating the treatment effect, namely, the extent to which norm-violating behaviours are judged differently depending on the gender of the non-complier. Next, we test whether the judgment depends on the gender of the respondent and, in particular, whether same-gender judgments are different from cross-gender ones. This is an important research question, as cross-gender effects have been shown to play a role by, e.g., Asmat et al. (2024). Finally, we analyse the impact of the gender of the enumerator in the field to infer the existence of experimenter demand effects.

Our main specification is as follows:

$$Index_i = \alpha_t + \beta\, Treatment_i + X_i\, \delta + \epsilon_i \tag{1}$$

where $Index_i$ is the sum of the responses to the three questions by subject i, and α_t are day fixed effects (from Tuesday to Friday). The pre-registered vector of socioeconomic controls X_i comprises a gender indicator (that takes value one if the respondent is a female), age, and

dummy variables that take value one if the respondent is a foreigner, if she/he self-reported to be in Lisbon specifically for the WYD, and if she/he has a bachelor degree. In addition, we also control for religious preferences using the centrality of religiosity scale (Huber and Huber, 2012). We tackle the second research question by interacting the treatment indicator with the gender of the respondent. Similarly, the third research question is investigated with an interaction term with the gender of the assistant.

In Table 2, we find that women were less morally judged than men. Taking into account an average index mean of 13.19 points, seeing a female name caused a reduction of around 4% in this score. When we analyze each of the individual items of the index, we observe that this gap is particularly evident when we consider the items "married person having sex with strangers" and "not picking up the kids from school because one is working until late". These effects are 7% and 2.7%, respectively. There is no difference in judgment for "lying to the boss".

Point estimates are very similar without and with a vector of control variables, lending further credibility to the idea that the randomization process successfully balanced the control and treatment groups. We highlight that men and women are equally judgmental, while older people seem to be, on average, more judgmental. In addition, a more religious person (proxied by the religiosity index) tends to judge more, on average, and we do not find differences in judgment for more educated respondents.

We also show that there are no significant heterogeneous effects depending on nationality (comparing Portuguese with foreigners), education (proxied by a graduate degree), and for older respondents (above the median age of 23 years) in Table A4 in the Appendix. The results are also robust when we include assistant fixed effects in Table A5 in the Appendix.

Next, we test whether there are significant differences in judgment depending on the respondents' gender by interacting the treatment with the female respondent dummy. In Table 3, we show that, in our sample, the average levels of moral judgment of women and men in the index are very similar. When looking at specific stories, men and women have

Table 2: Women were less morally judged than men

	Index		Lied to boss		Did not pick kids		Sex with stranger	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	-0.492***	-0.479***	-0.101	-0.093	-0.243***	-0.241***	-0.149**	-0.146**
	(0.141)	(0.138)	(0.070)	(0.069)	(0.077)	(0.075)	(0.062)	(0.061)
female respondent		-0.062		-0.185***		0.119		0.004
		(0.140)		(0.070)		(0.077)		(0.063)
age		0.043***		0.013***		0.029***		0.000
		(0.007)		(0.004)		(0.004)		(0.003)
in Lisbon for WYD		-0.227		-0.018		-0.149		-0.060
		(0.233)		(0.112)		(0.126)		(0.099)
religiosity index		0.582***		0.260***		0.099		0.222***
		(0.137)		(0.066)		(0.074)		(0.062)
foreigner		-0.123		-0.123		0.117		-0.116*
		(0.164)		(0.080)		(0.089)		(0.070)
graduate		-0.125		-0.076		0.007		-0.057
		(0.152)		(0.077)		(0.083)		(0.069)
Day Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
N	1,800	1,800	1,800	1,800	1,800	1,800	1,800	1,800
Mean Y	13.19	13.19	4.43	4.43	3.36	3.36	5.40	5.40

Notes: Robust standard errors in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

the same levels of judgment for having sex with strangers.

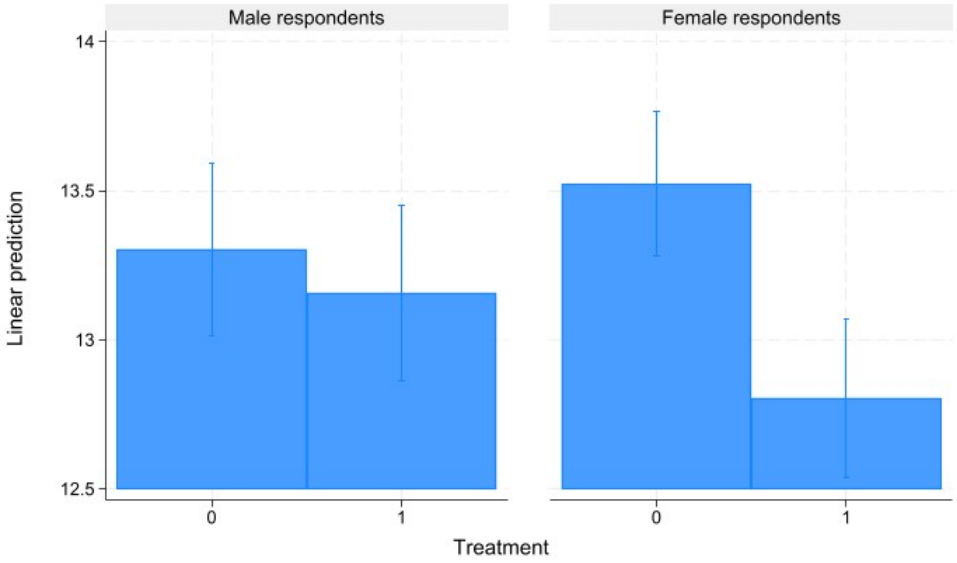
However, women respondents are less judgmental towards women than men, while men respondents do not discriminate by gender. When looking at specific stories, we find that this effect is mostly driven by women discriminating for not picking up the children from school, namely, they judge less involved mothers less harshly than less involved fathers. These effects are consistent with the growing literature on the child penalties related to gender and parenthood (Kleven et al., 2019, 2024).

The marginal effects are displayed in Figure 3. The results remain robust when we include a vector of assistant fixed effects as shown in Table A6 in the Appendix.

We then examine whether the gender of the assistant who implemented the survey had an impact on moral judgment in Table 4. We find no significant differences in outcomes related to the gender of the survey assistant, suggesting that the effects of the experimenter's demand may not be particularly relevant in this case (Zizzo, 2010; De Quidt et al., 2018, 2019).

To test whether our results are driven by more or less conservative respondents, we exploit heterogeneity in agreement with two statements that we designed to assess fundamentalism.

Figure 2: Marginal effects



Notes: 95% confidence intervals

Table 3: Women discriminated morally by gender while men did not

	Index		Lied to boss		Did not pick kids		Sex with stranger	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	-0.230	-0.146	-0.108	-0.081	-0.035	-0.001	-0.087	-0.064
	(0.214)	(0.210)	(0.106)	(0.105)	(0.118)	(0.116)	(0.098)	(0.097)
female respondent	0.199	0.219	-0.182*	-0.175*	0.310***	0.321***	0.071	0.073
	(0.196)	(0.194)	(0.098)	(0.098)	(0.110)	(0.109)	(0.087)	(0.087)
Treatment × female respondent	-0.450	-0.572**	0.021	-0.020	-0.364**	-0.412***	-0.107	-0.140
	(0.284)	(0.280)	(0.141)	(0.140)	(0.156)	(0.152)	(0.127)	(0.126)
age		0.043***		0.013***		0.029***		0.001
		(0.007)		(0.004)		(0.004)		(0.003)
in Lisbon for WYD		-0.236		-0.018		-0.156		-0.062
		(0.231)		(0.112)		(0.126)		(0.098)
religiosity index		0.593***		0.261***		0.107		0.225***
		(0.137)		(0.066)		(0.074)		(0.062)
foreigner		-0.115		-0.123		0.122		-0.114
		(0.163)		(0.080)		(0.089)		(0.070)
graduate		-0.123		-0.075		0.009		-0.056
		(0.152)		(0.077)		(0.083)		(0.069)
Day Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
N	1,800	1,800	1,800	1,800	1,800	1,800	1,800	1,800
Mean Y	13.19	13.19	4.43	4.43	3.36	3.36	5.40	5.40

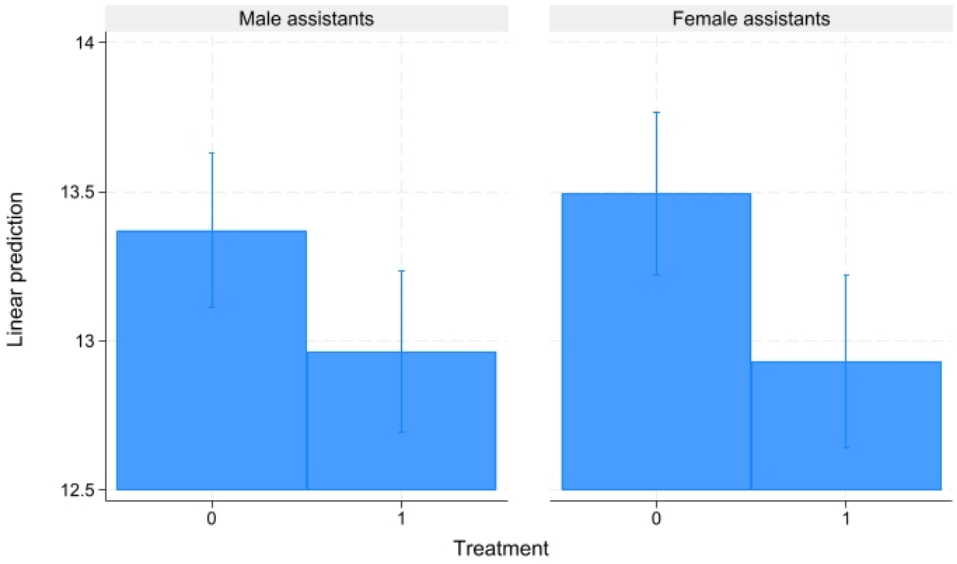
Notes: Robust standard errors in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: The gender of the assistant did not impact moral judgment

	Index		Lied to boss		Did not pick kids		Sex with stranger	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	-0.418**	-0.406**	-0.086	-0.070	-0.211**	-0.203*	-0.121	-0.133
	(0.193)	(0.190)	(0.097)	(0.096)	(0.107)	(0.104)	(0.085)	(0.084)
female assistant	0.123	0.124	0.036	0.054	0.118	0.128	-0.032	-0.059
	(0.194)	(0.193)	(0.098)	(0.097)	(0.109)	(0.109)	(0.086)	(0.086)
Treatment × female assistant	-0.158	-0.156	-0.032	-0.048	-0.071	-0.083	-0.056	-0.025
	(0.283)	(0.278)	(0.140)	(0.139)	(0.154)	(0.152)	(0.125)	(0.124)
female respondent		-0.069		-0.188***		0.111		0.009
		(0.141)		(0.070)		(0.077)		(0.063)
age		0.043***		0.013***		0.029***		0.000
		(0.007)		(0.004)		(0.004)		(0.003)
in Lisbon for WYD		-0.225		-0.017		-0.149		-0.058
		(0.233)		(0.112)		(0.127)		(0.099)
religiosity index		0.580***		0.259***		0.097		0.224***
		(0.137)		(0.066)		(0.074)		(0.062)
foreigner		-0.120		-0.122		0.120		-0.118*
		(0.164)		(0.080)		(0.090)		(0.071)
graduate		-0.130		-0.078		-0.000		-0.052
		(0.152)		(0.077)		(0.084)		(0.069)
Day Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
N	1,800	1,800	1,800	1,800	1,800	1,800	1,800	1,800
Mean Y	13.19	13.19	4.43	4.43	3.36	3.36	5.40	5.40

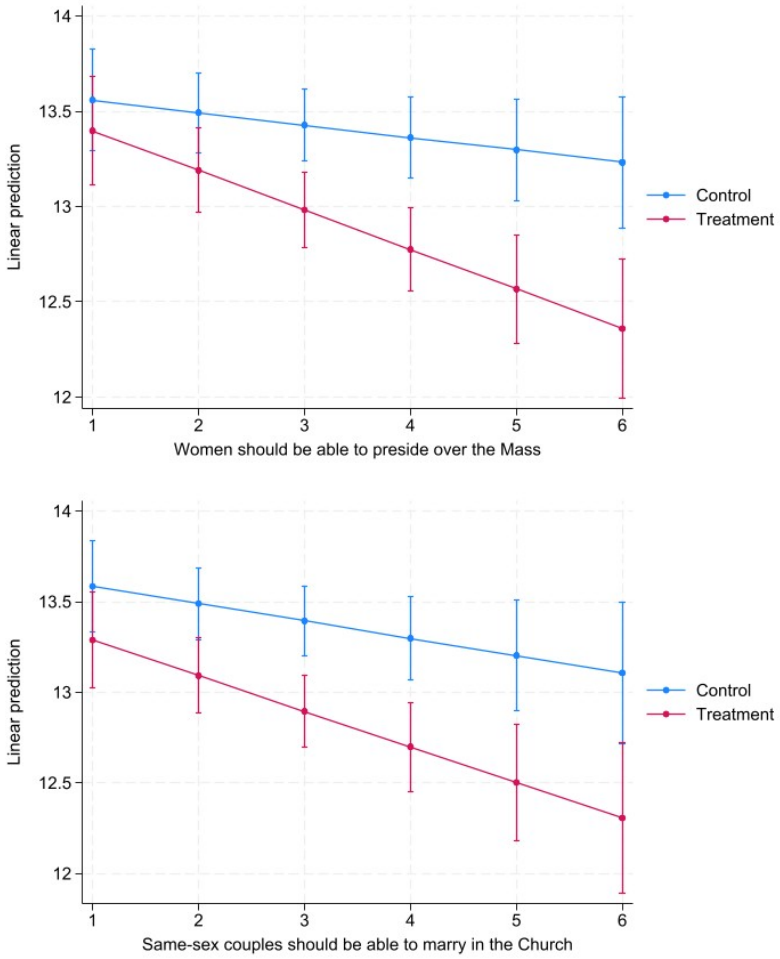
Notes: Robust standard errors in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 3: Marginal effects



Notes: 95% confidence intervals

Figure 4: Heterogeneity



Notes: 95% confidence intervals

The results are presented in Figure 4 and exhibit similar patterns in the two panels. In panel a), the statement is "women should be able to preside over the mass" while in panel b) the statement is "same-sex couples should be allowed to marry in the Church". We find that more conservative respondents (i.e., those with lower values in the x-axis) judge more than less conservative ones, but do not discriminate by gender. On the other hand, less conservative respondents tend to judge women (treatment) less than men (control).

5 Concluding remarks

Gender stereotypes, both descriptive and prescriptive, are pervasive, with the latter often implying penalties for norm violation. However, the determinants of gender prescriptive stereotypes (i.e., how males and females *should* behave) are not well studied. We provide the first field experiment dealing with prescriptive gender stereotyping on moral issues. Moreover, we do so in the context of the World Youth Day 2023 in Lisbon, a Catholic event that attracted around 2 million pilgrims to the city, during which moral judgments and gender stereotypes are likely to become salient.

We show that respondents exhibit high levels of judgment, with women being, on average, less judged than men. We also find that men do not discriminate according to gender, while women tend to be more lenient about other women's behaviour. These effects are driven by the parental care scenario. Importantly, more conservative respondents are more judgmental, but do not discriminate between genders, while less fundamentalist ones are less judgmental, but discriminate in favour of women. Moreover, the gender of the enumerator allows us to show that experimenter demand effects are not very relevant in this setting.

Our results have important policy implications. Moral judgments can affect jury decisions, hiring practices, or parental custody rulings. As these moral judgments are likely to affect individual preferences, they also can have an impact on education and professional choices, with implications for gender gaps. We also provide important insights about gender discrimination in the teachings of the Catholic church, that plays an important role in shaping moral preferences of millions of people around the world.

The construction and evaluation of the effectiveness of public policies to mitigate gender discrimination depends on the existence of discrimination in the family and religious environment, and this discrimination needs to be rigorously studied.

References

Asmat, R., Borowiecki, K., and Law, M. T. (2024). Competing for equality: Gender bias among juries in international piano competitions, 1890-2023. *Available at SSRN 4910483*.

Bassi, V. and Rasul, I. (2017). Persuasion: A case study of papal influences on fertility-related beliefs and behavior. *American Economic Journal: Applied Economics*, 9(4):250–302.

Becker, S. O., Rubin, J., and Woessmann, L. (2021). Religion in economic history: A survey. *The Handbook of Historical Economics*, pages 585–639.

Bertrand, M. and Duflo, E. (2017). Field experiments on discrimination. *Handbook of economic field experiments*, 1:309–393.

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2016). Stereotypes. *The Quarterly Journal of Economics*, 131(4):1753–1794.

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2019). Beliefs about gender. *American Economic Review*, 109(3):739–773.

Bursztyn, L., Cappelen, A. W., Tungodden, B., Voena, A., and Yanagizawa-Drott, D. H. (2023). How are gender norms perceived? Technical report, National Bureau of Economic Research.

Bursztyn, L., González, A. L., and Yanagizawa-Drott, D. (2020). Misperceived social norms: Women working outside the home in Saudi Arabia. *American economic review*, 110(10):2997–3029.

Bursztyn, L. and Yang, D. Y. (2022). Misperceptions about others. *Annual Review of Economics*, 14(1):425–452.

Casciaro, T. and Lobo, M. S. (2005). Competent jerks, lovable fools, and the formation of social networks. *Harvard Business Review*, 83(6):92–99.

Charness, G., Gneezy, U., and Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, 81(1):1–8.

Clingingsmith, D., Khwaja, A. I., and Kremer, M. (2009). Estimating the impact of the Hajj: religion and tolerance in Islam's global gathering. *The Quarterly Journal of Economics*, 124(3):1133–1170.

De Quidt, J., Haushofer, J., and Roth, C. (2018). Measuring and bounding experimenter demand. *American Economic Review*, 108(11):3266–3302.

De Quidt, J., Vesterlund, L., and Wilson, A. J. (2019). Experimenter demand effects. In *Handbook of research methods and applications in experimental economics*, pages 384–400. Edward Elgar Publishing.

Dhar, D., Jain, T., and Jayachandran, S. (2022). Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in India. *American Economic Review*, 112(3):899–927.

Ellemers, N. (2018). Gender stereotypes. *Annual Review of Psychology*, 69(1):275–298.

Farina, E. and Pathania, V. (2020). Papal visits and abortions: evidence from Italy. *Journal of Population Economics*, 33(3):795–837.

Glick, P. and Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, 70(3):491—-512.

Gundersen, S. J. and Beck, S. V. (2020). *Religion and Gender*, pages 1–22. Springer International Publishing, Cham.

Heilman, M. E., Caleo, S., and Manzi, F. (2024). Women at work: pathways from gender stereotypes to gender bias and discrimination. *Annual Review of Organizational Psychology and Organizational Behavior*, 11(1):165–192.

Heilman, M. E., Wallen, A. S., Fuchs, D., and Tamkins, M. M. (2004). Penalties for success: reactions to women who succeed at male gender-typed tasks. *Journal of Applied Psychology*, 89(3):416.

Huber, S. and Huber, O. W. (2012). The centrality of religiosity scale (CRS). *Religions*, 3(3):710–724.

Kleven, H., Landais, C., and Leite-Mariante, G. (2024). The child penalty atlas. *Review of Economic Studies*.

Kleven, H., Landais, C., and Søgaard, J. E. (2019). Children and gender inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, 11(4):181–209.

Niederle, M. and Vesterlund, L. (2011). Gender and competition. *Annual Review of Economics*, 3(1):601–630.

Rudman, L. A. and Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues*, 57(4):743–762.

Sinding Bentzen, J. (2019). Acts of god? religiosity and natural disasters across subnational world districts. *The Economic Journal*, 129(622):2295–2321.

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

Appendix

Table A1: Nationalities in our sample

Nationality	Freq.	Percent
portuguese	516	28.67
spanish	223	12.39
brazilian	146	8.11
french	137	7.61
american	116	6.44
mexican	77	4.28
italian	39	2.17
polish	35	1.94
british	30	1.67
colombian	28	1.56
german	26	1.44
australian	25	1.39
indian	22	1.22
angolan, canadian	21	1.17
irish	19	1.06
phillipines	18	1.00
peruvian	16	0.89
panamanian, salvadoran	14	0.78
chilean, puerto rican	13	0.72
south african	11	0.61
dominican, venezuelan	10	0.56
argentinian, czech	9	0.50
ecuadorian, guatemalan, singaporean	8	0.44
belgian, croatian	7	0.39
austrian, honduran	6	0.33
cape verdean, cuban, dutch, east timorese, equatorial guinea, filipino, indonesian, kenyan, uruguayan	5	0.28
equatorian, ghanaian, hungarian, mozambican	4	0.22
bolivian, costa rican, lebanese, malay, nigerian, slovakian, tanzanian, zimbabwean	3	0.17
belize, cameroonese, chinese, congolese, danish, finnish, jordanian, mauritian, romanian, serbian, swiss, vietnamese	2	0.11
azerbeijan, canada, greek, icelandic, iraqi, ivorian, japanese, lithuanian, maltese, myanmar, nepalese, nicaraguan, pakistani, paraguay, saint thomas, seychelles, slovenian, solomon islanders, south korean, thai, uae, ugandan, ukrainian	1	0.06
Total	1800	100

Table A2: Balance Table: Treatment

	Control	Treatment	Difference
female respondent	0.567	0.597	0.030
	(0.496)	(0.491)	(0.023)
female assistant	0.464	0.483	0.019
	(0.499)	(0.500)	(0.024)
age	27.317	27.255	-0.062
	(10.786)	(10.308)	(0.497)
in Lisbon for WYD	0.884	0.896	0.012
	(0.320)	(0.305)	(0.015)
religiosity index	4.540	4.542	0.003
	(0.509)	(0.546)	(0.025)
foreigner	0.719	0.707	-0.012
	(0.450)	(0.455)	(0.021)
graduate	0.653	0.664	0.011
	(0.476)	(0.473)	(0.022)
women should be able to preside over the Mass	3.092	3.173	0.081
	(2.037)	(2.057)	(0.097)
same-sex couples should be able to marry in the Church	2.680	2.740	0.060
	(1.982)	(1.986)	(0.094)
Observations	905	895	1,800

Table A3: Balance Table: Gender

	Male	Female	Difference
female assistant	0.432	0.503	0.072***
	(0.496)	(0.500)	(0.024)
age	27.046	27.458	0.412
	(9.824)	(11.041)	(0.504)
in Lisbon for WYD	0.908	0.877	-0.032**
	(0.289)	(0.329)	(0.015)
religiosity index	4.517	4.558	0.041*
	(0.563)	(0.500)	(0.025)
foreigner	0.726	0.704	-0.023
	(0.446)	(0.457)	(0.022)
graduate	0.640	0.671	0.031
	(0.480)	(0.470)	(0.023)
women should be able to preside over the Mass	3.124	3.139	0.015
	(2.052)	(2.045)	(0.098)
same-sex couples should be able to marry in the Church	2.445	2.900	0.455***
	(1.882)	(2.033)	(0.094)
Observations	753	1,047	1,800

Table A4: Heterogeneity Table 2

	(1) Index	(2) Index	(3) Index
Treatment	-0.458*	-0.450**	-0.580***
	(0.239)	(0.228)	(0.176)
Treatment × foreigner	-0.030		
	(0.292)		
Treatment × graduate		-0.045	
		(0.287)	
Treatment × above median age			0.197
			(0.239)
female_respondent	-0.061	-0.061	-0.060
	(0.140)	(0.140)	(0.140)
age	0.043***	0.043***	0.040***
	(0.007)	(0.007)	(0.008)
in Lisbon for WYD	-0.228	-0.225	-0.227
	(0.232)	(0.233)	(0.233)
religiosity index	0.582***	0.582***	0.582***
	(0.137)	(0.137)	(0.137)
foreigner	-0.107	-0.123	-0.131
	(0.210)	(0.164)	(0.164)
graduate	-0.125	-0.103	-0.157
	(0.152)	(0.204)	(0.157)
Day Fixed Effects	Yes	Yes	Yes
N	1,800	1,800	1,800
Mean Y	13.19	13.19	13.19

Table A5: Robustness Table 2: including Assistant FE

	Index		Lied to boss		Did not pick kids		Sex with stranger	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	-0.486***	-0.472***	-0.106	-0.097	-0.245***	-0.242***	-0.135**	-0.132**
	(0.142)	(0.139)	(0.070)	(0.069)	(0.078)	(0.076)	(0.062)	(0.062)
female respondent		-0.022		-0.168**		0.122		0.024
		(0.143)		(0.071)		(0.078)		(0.064)
age		0.043***		0.013***		0.030***		0.000
		(0.008)		(0.004)		(0.004)		(0.003)
in Lisbon for WYD		-0.266		-0.030		-0.168		-0.068
		(0.236)		(0.113)		(0.127)		(0.100)
religiosity index		0.619***		0.264***		0.106		0.249***
		(0.139)		(0.067)		(0.075)		(0.064)
foreigner		-0.072		-0.122		0.159*		-0.109
		(0.168)		(0.082)		(0.093)		(0.075)
graduate		-0.120		-0.091		0.028		-0.058
		(0.155)		(0.080)		(0.086)		(0.070)
Day Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
Assistant Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
N	1,797	1,797	1,797	1,797	1,797	1,797	1,797	1,797
Mean Y	13.19	13.19	4.43	4.43	3.36	3.36	5.40	5.40

Notes: Robust standard errors in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Robustness Table 3: including Assistant FE

	Index		Lied to boss		Did not pick kids		Sex with stranger	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	-0.199	-0.112	-0.110	-0.082	-0.019	0.015	-0.071	-0.045
	(0.213)	(0.209)	(0.105)	(0.105)	(0.119)	(0.116)	(0.096)	(0.096)
female respondent	0.265	0.278	-0.161	-0.156	0.329***	0.337***	0.096	0.097
	(0.198)	(0.197)	(0.099)	(0.099)	(0.111)	(0.110)	(0.088)	(0.088)
Treatment × female respondent	-0.492*	-0.615**	0.014	-0.025	-0.394**	-0.440***	-0.112	-0.150
	(0.283)	(0.280)	(0.141)	(0.140)	(0.156)	(0.152)	(0.125)	(0.125)
age		0.043***		0.013***		0.030***		0.001
		(0.008)		(0.004)		(0.004)		(0.003)
in Lisbon for WYD		-0.273		-0.030		-0.173		-0.070
		(0.234)		(0.113)		(0.127)		(0.099)
religiosity index		0.632***		0.264***		0.115		0.253***
		(0.139)		(0.067)		(0.075)		(0.063)
foreigner		-0.064		-0.122		0.165*		-0.107
		(0.168)		(0.083)		(0.092)		(0.075)
graduate		-0.118		-0.091		0.030		-0.057
		(0.155)		(0.080)		(0.086)		(0.070)
Day Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
Assistant Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
N	1,797	1,797	1,797	1,797	1,797	1,797	1,797	1,797
Mean Y	13.19	13.19	4.43	4.43	3.36	3.36	5.40	5.40

Notes: Robust standard errors in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.