# Gendered language and gendered violence

ABSTRACT. This study establishes the influence of sex-based grammatical gender on gendered violence. We demonstrate a statistically significant relationship between speaking a gendered language and the incidence of intimate partner violence in a cross-section of countries. Motivated by this evidence, we conduct an individual-level analysis of the effect of speaking a gendered language on beliefs about the justifiability of intimate partner violence, controlling for a wide variety of individual level socioeconomic characteristics as well as country, religion, language family and ethnicity fixed effects. Speaking a gendered language is associated with the belief that intimate partner violence is justifiable. Our results are consistent with complementarity between the cultural and cognitive effects of language on the attitudes to intimate partner violence.

*JEL classification:* J12, J16, Z13, D83, D91.

*Keywords: gender-based violence, intimate partner violence, sex-based grammar, gendered language, gender norms.* 

#### 1 Introduction

Intimate Partner Violence (IPV) is pervasive. According to estimates by the World Health Organization (2018), 26% of women globally have been subjected to physical and/or sexual violence by their husband or male partner at least once in their lifetime. IPV has immense negative consequences for the health and well-being of survivors (Campbell, 2002; Devries et al., 2013; Beleche, 2019) and their children (Aizer, 2011; Currie et al., 2018; Anderberg & Moroni, 2020; Bharati et al., 2024). Moreover, IPV imposes significant social and economic costs on societies due to the lost income and decreased productivity of survivors and their families, expenditures for the provision of services, and negative impacts on future human capital formation (Duvvury et al., 2013; Borker, 2017). For example, the annual costs of IPV are estimated to be \$5.8 billion in the United States of America and \$1.16 billion in Canada (UN Women, 2016). Yet, our understanding of the factors which contribute to the prevalence of IPV remains insufficient. This paper aims to partially fill the gap by offering a novel explanation to the literature: sex-based gender marking in language.

Different languages vary strongly with regards to grammatical gender. In some languages gender marking is ubiquitous, in others it is entirely absent (Corbett, 2013a). In many languages biological sex forms the semantic core of the gender system. Sex-based linguistic gender can be related to attitudes and behaviours concerning gendered violence through two different mechanisms, a cultural and a cognitive one (or through a combination of the two).

First, according to the cultural channel, gendered language may be associated with IPV because such language serves as a marker for gender-biased social norms. As argued by Galor et al. (2018), language structures may evolve in a manner that complements existing cultural values and, indeed, they find evidence that sex-based gender systems are more common in languages that developed in regions conducive to a gender gap in agricultural productivity. Such languages force speakers to pay close attention to gender distinctions and hierarchies. In the case of linguistic gender, the cultural mechanism suggests that in places that assign a high significance to gender related social norms, there is a higher probability

that gender distinctions are reflected in language. In the current context, this means that gendered language may not have a causal relationship to IPV.

Alternatively, according to the cognitive mechanism, linguistic features may have direct causal effect on our cognition and behavior, an idea that is commonly referred to as the *Linguistic Relativity Hypothesis*. In the case of linguistic gender the cognitive mechanism can be manifested in the following form. We structure reality, including the social world, using mental *schemata* (Rumelhart, 1980; Tajfel, 1981; Turner et al., 1987; Fiske & Taylor, 1991). Gendered language leads to an increased activation of gender schemata in the minds of speakers, and thereby raises the salience of gender distinctions and the significance of gender norms in the minds of speakers. As a consequence, existing gender norms, which may legitimise IPV, may get reinforced and acquire a greater influence on speakers of a gendered language. The cognitive channel suggests that sex-based linguistic gender has an independent causal effect on individuals' cognition. It is of course also possible that both channels are operative. Most likely, there is a complex and reciprocal causal interaction between linguistic and cultural features.

We empirically examine the relationship between gendered language and IPV in two ways. First, using cross-country data, we consider the relationship between gendered language and the reported incidence of IPV. Controlling for a key set of economic, religious and institutional variables, we find that the share of women having experienced IPV is 8 percentage points larger in countries where the language spoken by the majority of the population is gendered. While we make no claim to having identified a causal relationship, the strong association between gendered language and IPV in the international data serves to motivate our primary empirical investigation, which uses individual level data to consider an important mechanism that may underly this relationship, the link between speaking a gendered language and expressing a belief in the justifiability of IPV.

Based on a global data collection of nationally representative individual-level surveys, we show that speaking a gendered language is statistically significantly positively associated with the probability of considering wife beating justifiable. Relative to individuals whose language does not have a sex-based gender system, speaking a gendered language is associated with 7.5 percentage point increase in the probability of justifying wife beating. To

mitigate the confounding effects of geographic, cultural, historical, institutional and socioeconomic conditions, we control for country, wave, language family, religion and ethnicity fixed effects in our primary specification.

Our key finding is robust to excluding high-IPV regions of the world and speakers of global languages from our sample. In addition, we conduct a number of placebo tests, finding that speaking a gendered language is not empirically associated with beliefs in the justifiability of non-gendered forms of violence, including political violence and violence against children and others. The results of these tests support the theoretical mechanism we propose, which is specific to gender.

In spite of the extensive set of controls we employ, it is still possible that our result is spurious. In particular, it may be that the empirical association between gendered language and beliefs about the justifiability of wife beating reflects the influence of unobserved dimensions of culture rather than the causal influence of language itself. To address this concern, we employ a quasi-experimental design, whereby we explore whether for speakers of gendered languages, using a non-gendered interview language de-activates attitudes justifying gendered violence; and conversely, whether for speakers of non-gendered languages, using a gendered interview language activates these attitudes. We find that for individuals who speak a gendered language at home, using a genderless language to elicit views on gendered violence is associated with a reduced probability of tolerating such violence—additional evidence that suggests that the absence of a sex-based linguistic gender system might, through the reduced activation of gender schemata, also reduce the influence of social norms supportive of gendered violence.

Our finding that the cognitive channel may play a role in influencing beliefs about the justifiability of wife beating has important implications for the efficacy of language-based policy interventions. There has been a significant push across a variety of countries to adopt less gendered language structures in an attempt to address sexism and gender inequality. This includes the use of "they" as a non-gendered form of the third person singular in English (Baranowski, 2002), replacement of the pronouns "him" and "her" with an artificial genderless pronoun "hen" in schools in Sweden (Tagliabue, 2012), a move to revise the treatment of collective nouns as masculine in French (McCoubrey, 2017), and the use of "-e"

as a gender-neutral noun ending in Spanish (Politi, 2020). These and other language-based interventions are likely to have a greater influence on attitudes concerning IPV if the link between gendered language and beliefs is primarily cognitive in nature. Alternatively, if the empirical relationship between gendered language and beliefs about IPV arises because gendered language is correlated with unobserved dimensions of gender-biased culture, then the impact of language-based interventions may be limited.

This study contributes to the emerging body of work that seeks to identify the drivers of violence against women. Existing studies highlight the significance of historical circumstances that have shaped social norms that legitimize such violence, including the legacies of conflict (La Mattina, 2017; Mavisakalyan & Minasyan, 2023), historical family structures (Tur-Prats, 2019; Beltrán Tapia & Gallego-Martínez, 2020), and socio-economic hierarchies (Leyaro et al., 2017; Alesina et al., 2021). Contemporary informal norms around marriage and post-marital residence continue to influence violence experienced by women (Jacoby & Mansuri, 2010; Jayachandran, 2015; Khalil & Mookerjee, 2019). Formal institutions, like the features of inheritance and divorce laws and law-and-order institutions also affect such violence (Amaral, 2017; Amaral et al., 2021; García-Ramos, 2021). There is evidence to suggest that in some settings IPV declines in response to women's empowerment (Aizer, 2010; Hidrobo et al., 2016; Mavisakalyan & Rammohan, 2021). However, in other contexts, especially those characterised by conservative and rigid gender role norms, backlash against women's economic empowerment may increase aggression against women (Hidrobo & Fernald, 2013; Heath, 2014; Zhang & Breunig, 2021). Violence against women is also affected by systemic shocks like political unrest (Bargain et al., 2019), structural changes in the economy (Kotsadam et al., 2017), extreme weather events (Cools et al., 2020) and during periods of decreased mobility as in the current COVID-19 pandemic (United Nations, 2020; Arenas-Arroyo et al., 2021; Berniell & Facchini, 2021). We aim to contribute to a better understanding of the causes of IPV norms and the prevalence of IPV by studying gender in language as a possible source.

That gendered language might promote gender inequalities has been the subject of feminist scholarship for several decades (see Saul (2010) for an overview). Only recently has empirical evidence on the connection between gender inequalities and gendered language started to emerge. Existing studies show that gendered language is associated with more unequal outcomes in education (Davis & Reynolds, 2018; Jakiela & Ozier, 2018; Galor et al., 2020), employment (Mavisakalyan, 2015; van der Velde et al., 2015; Jakiela & Ozier, 2018) and corporate and political leadership (Santacreu-Vasut et al., 2013, 2014), as well as less liberalized attitudes toward gender equality (Pérez & Tavits, 2019).<sup>1</sup> Of particular relevance to the current study are the papers that link gendered language with gender inequalities in the household, reflected in the allocation of household production tasks on the basis of sex (Hicks et al., 2015) and labor market engagement of women within the context of household decision-making (Gay et al., 2018). We extend this literature by focusing on IPV norms and the prevalence of IPV as outcomes of speaking a gendered language.

As mentioned above, the idea that language has an independent influence on thought and behavior is known as the Linguistic Relativity Hypothesis (LRH) (for overviews see e.g. Gumperz & Levinson, 1996; Lucy, 1997; Boroditsky, 2003; Casasanto, 2015). A strong version of the LRH, which holds that language determines thought (Whorf, 1956), was initially rejected in linguistics (see e.g. Pinker, 1994; Gleitman & Papafragou, 2013). However, more recently the LRH has experienced a revival in psychology and cognitive science, and there is now a substantial body of research supporting a more moderate version of the LRH, which holds that language influences thought (see e.g. Lucy, 1996; Slobin, 2003; Boroditsky et al., 2003; Oh, 2003; Levinson & Wilkins, 2006; Kay & Regier, 2006; Winawer et al., 2007). Insofar as the connection between linguistic gender and IPV observed here follows the above-described cognitive channel, our study can be seen as a further contribution to this development.

The following section presents a conceptual framework underlying the proposed link between gendered language and IPV. Section 3 takes a first look at the link through a descriptive analysis of cross-country data. Section 4 presents the empirical approach and data underlying the main part of our analysis. The results of this analysis appear in Section 5. Finally, Section 6 concludes the paper and discusses its implications.

<sup>&</sup>lt;sup>1</sup> For reviews of this literature see Mavisakalyan & Weber (2017); Ginsburgh & Weber (2020).

## 2 Background

## 2.1 Grammatical Gender

Different languages vary strongly with regard to grammatical gender. Some languages have near-ubiquitous gender marking, other languages completely lack grammatical gender (Corbett, 2013a). Among the languages that do have gender, the number of grammatical genders can vary from 5 or more genders to only two. Another important dimension of variation concerns the semantic basis for gender distinctions, i.e. which real world feature gender is tracking. In most cases, the semantic core of gender is biological sex, but grammatical gender can also track non-sex based characteristics, such as the human/non-human distinction or the contrast between animate and inanimate objects (Corbett, 2013c). Structurally, grammatical gender is based on a system of *agreement* between groups of nouns and other expressions in a sentence (Corbett, 1991). Consider e.g. the following Russian sentences (Corbett, 2013a):

- Žurnal ležal na stole.
   magazine lay-M on table.
   The magazine lay on the table.
- (2) Kniga ležal-a na stole.book lay-F on table.*The book lay on the table.*
- (3) Pis´ mo ležal-o na stole.letter lay-N on table.*The letter lay on the table.*

Here, we find different verbs forms in (1), (2), and (3), depending on the different nouns in subject position, Žurnal, Kniga, Pis´ mo. For instance, since Pis´ mo has neutral gender, the verb needs to end in -o, also indicating neutral gender. Other forms of gender agreement may demand that e.g. adjectives or articles agree with the relevant noun.

As just mentioned, biological sex is the semantic core of the gender system in many languages. In French the noun for 'woman' has female grammatical gender ('*la* femme') and the noun for 'man' has male grammatical gender ('*le* homme'). But these sex based distinction are often projected far beyond the biological realm onto non-biological objects as well, e.g. 'la lune' (the moon) has female grammatical gender and 'le soleil' (the sun) has male grammatical gender. Following an established linguistic source (Dryer & Haspelmath, 2013), we distinguish between languages with a sex-based gender system (e.g. Spanish, Russian) on the one hand, and languages that do not have a sex-based gender system (e.g. Zulu, Swahili) or that lack gender altogether (e.g. Finnish, Persian) on the other (Corbett, 2013b). For simplicity, we refer to languages with a sex-based gender as 'gendered' languages and languages that lack a gender system or whose gender system is not sex-based both as 'genderless'. Our analysis explores whether speaking a gendered language is associated with higher acceptance of IPV legitimizing norms and a higher prevalence of IPV.

#### 2.2 Gendered language and gender-related social norms

Linguistic gender can be related to social norms concerning IPV through two broad mechanisms, a cultural and a cognitive one. Consider first the cultural mechanism. Some societies assign a higher importance to gender norms than others. In such societies, gender distinctions play a more significant socio-cultural role. As a result, these distinctions are more likely to find a linguistic manifestation. That is, language structures tend to evolve such that they complement cultural values. For instance, Galor et al. (2018) provide evidence that geographic features linked to cultural norms are also coded in language structures. There is then a greater likelihood to find a grammaticalised sex-based gender system in the language spoken in societies that place a higher emphasis on gender distinctions.

Second, sex-based linguistic gender may directly affect speakers' cognition and behaviour. We structure the natural and the social world using mental *schemata* (Rumelhart, 1980; Tajfel, 1981; Turner et al., 1987; Fiske & Taylor, 1991; Minsky, 1974). In the social domain, these schemata organise how we perceive others and ourselves. They also play an action-guiding role in the form of behavioral *scripts* (Schank & Abelson, 1977). A large amount of our behavior in social situations is based on such routines and social scripts. Gender schemata play a particularly important role in the social domain. From an early age on, children categorize people in their environment according to gender schemata (Martin & Halverson Jr, 1981; Ruble et al., 2006; Martinez et al., 2020). Schemata and scripts often incorporate *prototypes*, i.e. typical or ideal exemplars of the relevant category (Rosch, 1975, 1978). For instance, the schema for *wife* is connected to a prototypical wife which, depending on the context, may be viewed as someone who is obedient, does the cooking, cares for the children, is faithful, sexually available to the husband, etc. The wife-schema is related to a script that regulates how a husband should behave in relation to her, i.e. which behaviors are permissible or mandated. Consequently, schemata and scripts may encapsulate, through the associated norms and prototypes, damaging patriarchal social hierarchies and dominance structures.

The primary effect of speaking a language with a sex-based gender system is to increase the psychological activation of sex-based gender schemata and scripts. This may reinforce existing traditional gender norms, which often legitimise IPV, and increase their influence on speakers' cognition and behavior by facilitating the categorization of social situations in gendered terms. It may further raise the prominence of gender proto- or ideal-types, such as the ideal of the good wife.

While the cultural channel regards linguistic structures as causally inert manifestations of culture, the cognitive channel sees (certain) linguistic structures as having an independent effect on thought. Inevitably, linguistic structures arise in a certain cultural context. But once in place, they may acquire an autonomous causal influence on cognition. Below is a schematic way of visualizing the two mechanisms.



To conclude, sex-based linguistic gender is potentially linked to IPV related attitudes in two ways: through a cultural and/or through a cognitive pathway. Our picture is in line with *gendered resource theory* (Atkinson et al., 2005), according to which gender ideology and gender norms play an important role in understanding gendered domestic violence.

## 3 A first look: Gendered language and the incidence of IPV across countries

In this section, we provide evidence on the relationship between speaking a language with a sex-based gender system and the occurrence of IPV in a cross-section of countries. Our dependent variable, sourced from Georgetown Institute for Women, Peace and Security and Peace Research Institute Oslo (2019), captures the percentage of women who experienced physical or sexual violence committed by their intimate partner in the previous 12 months—a measure that is drawn from prevalence surveys conducted over the period 2000– 2017 (data reported in the series are the most recent available in the period specified). The presence of a sex-based gender system is defined with reference to the most widely-spoken language in a country taken from Alesina et al. (2003). We treat a language as gendered if it employs a sex-based system as defined in Corbett (2013b) and taken from the World Atlas of Language Structures (WALS) (Dryer & Haspelmath, 2013).<sup>2</sup>

We study whether cases of IPV are more prevalent in countries where the majority speaks a gendered language in a regression framework that includes the following sets of controls in a sequential fashion: (i) GDP per capita and its squared term (source: World Bank (2022)); (ii) lack of political rights and civil liberties (source: Freedom House (2022)); (iii) legal origins (source: La Porta et al. (2008)); (iv) religious make-up of the population (source: Barro (2003)). All models include language family and continent fixed effects. The results reported in columns (1)-(6) of Table 1 show that having a gendered language is associated with a higher share of women who have experienced IPV.<sup>3</sup> Based on our most extensive specification reported in column (6), moving from a genderless to a gendered language is associated with an 8.5 percentage point increase in a country's share of women exposed to IPV.

Our approach to assigning the marker of gendered language to the most spoken language in the country may be problematic, as some countries are multilingual. In some cases, the most spoken language in a country may not be the native language but rather the language

<sup>&</sup>lt;sup>2</sup>WALS is a large authoritative source of structural linguistic features compiled by linguists which we follow for both definition and classification of languages into gendered- and non-gendered groups in line with the recent studies by social scientists that have solely relied on WALS for definition and characterisation of linguistic features (e.g. Galor et al., 2020; Mavisakalyan et al., 2022). This is important in view of the strong criticism that studies using alternative approaches to coding, characterisation and sourcing of data on linguistic features (e.g. Chen, 2013) have received by linguists (e.g. Dahl, 2013; Roberts & Winters, 2013).

<sup>&</sup>lt;sup>3</sup>WALS provides on average classification for around 400 languages under each linguistic feature (classification of sex-based and non-sex-based gender systems is provided for 257 contemporary languages), however, as Comrie et al. (2013) note, this is still a relatively small share of the world's languages. In our case, while information on IPV prevalence is available for 169 countries, information on the sex-based and non-sex-based gender systems of a country's most widely spoken language is available for only 92 of these countries which explains the sample sizes employed in the analysis presented in Table 1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Depen	NDENT VAI	RIABLE: %	WOMEN E	XPERIENC	ed IPV
	All	All	All	All	All	All	w/o top 20
							diverse countries
Sex-based gender	9.292*	8.628**	9.515**	8.803**	7.388**	8.482**	7.655*
	(4.726)	(3.573)	(3.635)	(3.819)	(3.670)	(4.087)	(3.927)
LN GDP PER CAPITA		-9.227	-8.273	-4.519	-9.384	-4.563	-35.022
		(12.902)	(12.326)	(13.575)	(13.577)	(14.546)	(25.065)
LN GDP PER CAPITA SQUARED		0.271	0.295	0.016	0.287	0.092	1.655
		(0.667)	(0.624)	(0.686)	(0.704)	(0.741)	(1.316)
LACK OF POLITICAL RIGHTS			-2.038			-2.143	-3.637
			(1.748)			(2.160)	(3.006)
LACK OF CIVIL LIBERTIES			3.904*			4.037	5.606
			(2.090)			(2.536)	(3.461)
Common law origin			. ,	0.616		-0.721	-2.045
				(1.739)		(2.205)	(2.289)
CIVIL LAW ORIGIN				0.292		-1.421	-0.328
				(1.517)		(2.328)	(2.218)
Socialist origin				-3.250		-12.110	-8.673
				(3.290)		(12.577)	(12.880)
CATHOLIC SHARE				. ,	0.088*	0.103**	0.014
					(0.046)	(0.043)	(0.048)
PROTESTANT SHARE					0.076	0.095*	-0.014
					(0.050)	(0.056)	(0.056)
Orthodox share					0.063	0.069	-0.033
					(0.055)	(0.053)	(0.061)
MUSLIM SHARE					0.035	0.046	0.012
					(0.048)	(0.051)	(0.096)
Hindu share					0.156**	0.127	-0.004
					(0.077)	(0.084)	(0.125)
BUDDHIST SHARE					-0.009	0.131	-0.013
					(0.063)	(0.198)	(0.213)
Language family dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Continent dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.564	0.657	0.688	0.643	0.679	0.696	0.703
Mean of dependent	13.012	12.883	12.883	12.611	12.883	12.611	10.225
variable							
Ν	92	89	89	88	89	88	68

Table 1: Gendered language and IPV across countries — OLS coefficients

Note.— Standard errors in parentheses. \*Denotes significance at 10 percent; \*\*at 5 percent; \*\*\*at 1 percent levels. Column (7) excludes the top 20 linguistically fractionalized countries based on Alesina et al. (2003), including Uganda, Tanzania, Togo, Cameroon, Kenya, South Africa, Chad, Nigeria, Philippines, Central African Republic, Ethiopia, India, Cote d'Ivoire, Gabon, Guinea, Iran, Sudan, Ghana, Niger and Thailand.

of the former coloniser. This will introduce a measurement error in the definition of SEX-BASED GENDER. A possible solution to the problem would be to use a continuous measure of the share of individuals who speak gendered languages in each country. While data on the number of different language speakers by country is available from the Ethnologue (Eberhard & Fennig, 2024), the overlap of the languages with those in WALS is highly incomplete for most countries in the sample, making this data of limited use for our study. Still, we engage with the issue of the measurement error in SEX-BASED GENDER due to the presence of multilingual countries in the sample, by excluding the top 20 linguistically fractionalized countries based on Alesina et al. (2003) from the sample.<sup>4</sup> The results reported in column (7) of Table 1 show that the significant positive coefficient on SEX-BASED GENDER is retained within the remaining relatively less linguistically fractionalised sub-sample of countries.

We do not consider these findings evidence of a causal relationship between gendered language and IPV. The low frequency of cross-country data precludes investigating the full range of plausible confounding factors that may be at work here and as a result, we have concerns over the role of omitted variables and especially the complications posed by the correlation between gendered language and unobserved dimensions of patriarchal culture. Nonetheless, the fact that gendered language is correlated with the incidence of IPV in the cross-country data helps to motivate our search for a plausible causal mechanism linking speaking a gendered language and attitudes on the justifiability of IPV.

As Heise & Kotsadam (2015) show, the presence of norms that justify IPV is a significant predictor of IPV occurrence among individuals—a finding that can also be seen in Figure 1 where we plot % WOMEN EXPERIENCED IPV against % BELIEVE WIFE BEATING JUSTIFI-ABLE. The second variable comes from the World Values Surveys, a collection of nationally representative surveys on social, political, economic, religious and cultural values of people. Conducted since 1981 every 5 years, these surveys are a rich source of information on values in over 120 societies around the world. Waves 5–7 of WVS conducted over the period between 2005–2020 asked individuals to make a judgement on the extent of justifiability of physical IPV on a scale from 1 (never justifiable) to 10 (always justifiable). Specifically, participants were asked whether it was justifiable for a man to beat his wife. Our analysis distinguishes between individuals who believe such violence is never justifiable (0) from those who allow for some degree of justifiability of violence. We take the country-level averages of this variable for the purposes of the analysis presented in Figure 1.

Having established the reduced form relationship between gendered language and prevalence of IPV in a cross section of countries, we focus on studying the underlying mechanism in the remainder of the paper. Specifically, we conduct an individual level analysis of the

<sup>&</sup>lt;sup>4</sup>This includes Uganda, Tanzania, Togo, Cameroon, Kenya, South Africa, Chad, Nigeria, Philippines, Central African Republic, Ethiopia, India, Cote d'Ivoire, Gabon, Guinea, Iran, Sudan, Ghana, Niger and Thailand.



Figure 1: Relationship between IPV beliefs and incidence

Note.— % women experienced IPV is the percentage of women who experienced physical or sexual violence committed by their partner in the previous 12 months (Georgetown Institute for Women, Peace and Security and Peace Research Institute Oslo, 2019). % belief wife beating justifiable is the share of individuals who believe violence can be justified among all individuals surveyed in the World Values Survey in a country.

relationship between speaking a gendered language and finding IPV justifiable in a setup that allows us to account for a wide range of plausible confounding factors. We discuss this next.

## 4 Empirical approach and data

#### 4.1 Empirical model

The main part of our analysis focuses on individuals, linking the grammatical structure of their language with their beliefs on the justifiability of IPV. To that end, we estimate an equation of the following general form:

$$IPVJustifiability_{ictl}^* = \beta GenderedLanguage_{ictl} + \delta \mathbf{X_{ictl}}' + \varepsilon_{ictl}$$
(1)

where *IPVJustifiability*<sup>\*</sup><sub>ictl</sub> indicates whether the individual *i* in country *c* at time *t* who speaks a language *l* finds IPV justifiable or not.  $X_{ictl}$  is a set of characteristics that account for individual socio-economic and cultural background *i*, country *c*, survey wave *t*, and language family *l*, and  $\varepsilon_{ictl}$  is the error term.

We assume that the observed IPV justifiability belief  $IPV Justifiability_{ictl}$  relates to latent propensity through the criterion  $IPV Justifiability_{ictl} = 1(IPV Justifiability_{ictl}^* \ge 0)$ , so that the probability of holding an IPV-justifying belief under an assumption of normality for  $\varepsilon_{ictl}$ becomes a standard probit model as follows:

$$Pr(IPVJustifiability_{ictl} = 1 | GenderedLanguage_{ictl}, \mathbf{X_{ictl}}) = \Phi(\beta GenderedLanguage_{ictl} + \delta \mathbf{X_{ictl}}')$$
(2)

The presence of  $X_{ictl}$  ensures that the comparisons are drawn amongst individuals who are similar in a number of significant ways, including the characteristics of their location and their own cultural origin, both plausibly linked to the nature of beliefs and languages possessed by individuals. In our baseline regression, we control for country, wave and language family fixed effects, and we gradually expand our set of controls to include a range of individual level demographic and socio-economic characteristics, and religion and ethnicity fixed effects. We control for language families in all models, following Roberts et al. (2015), to account for geographical and historical relatedness of languages.

## 4.2 Data

To analyze the link between gendered language and beliefs on the justifiability of violence, we use waves 5–7 of the World Values Survey conducted over the period between 2005–2020 described earlier. In these waves, the individuals were asked to make a judgement on the extent of justifiability of physical IPV on a scale from 1 (never justifiable) to 10 (always justifiable). Our analysis distinguishes between individuals who believe it is never justifiable for a man to beat his wife (0) from those who allow for some degree of justifiability of violence.

To analyze how individuals' attitudes on the justifiability of IPV is related to the structure of their language, we use information on the language spoken at home. This variable it-self has been included in the surveys since Wave 3 and is available for the period 2005–2020

## Figure 2: Global distribution of speakers of languages that have sex-based gender and prevalence of justifiability of wife abuse



Source: World Bank Official Boundaries

studied. As noted earlier, the information on grammatical gender systems is from WALS online (Dryer & Haspelmath, 2013), an authoritative and widely used source of comparative linguistic information. This data is matched with information on the language an individual speaks at home, available for WVS Waves 3–9, to construct an individual-level variable SEX-BASED GENDER, an indicator whose value is one in case the language spoken at home has sex-based gender system (following the definition and classification used by linguists (Corbett, 2013b)). We construct a second sex-based language variable based on the language in which the survey is administered.

Note.— SEX-BASED GENDER is a binary variable for the presence of sex-based gender system in an individual's language defined following Corbett (2013b). It is averaged over all individuals surveyed in the World Values Survey in a country. Lighter shades indicate higher shares of SEX-BASED GENDER language speakers in a country. Countries with missing values are in grey. WIFE BEATING JUSTIFIABLE is a binary variable that distinguishes between individuals who believe it is never justifiable for a man to beat his wife (0) from those who allow for some degree of justifiability of violence. Green circles capture countrywide averages of WIFE BEATING JUSTIFIABLE with larger circles indicating higher prevalence of wife beating justifiability in a country.

The global distribution of speakers of languages with a sex-based gender systems is presented in Figure 2 (based on an unconstrained sample). As seen on the map, large parts of the world are coloured in either very light or very dark colours corresponding to the high degree of prevalence of gendered or genderless speakers in many countries. Indeed, the majority of countries in the sample have little variation in gendered language, since they have a single dominant language. Still, there are countries where we observe potentially significant within-county language differences which likely provide the majority of identification in our within-country regressions.<sup>5</sup> Based on the map, the average justifiability of wife beating varies across different parts of the world. However, based on these countrylevel averages, no clear patterns emerges. Our analysis is based on exploiting the withincountry differences in the structures of languages spoken by individuals to study the beliefs of justifiability of wife beating.

In various specifications, our analysis controls for a range of demographic and socioeconomic characteristics, including individuals' gender, age, family status, education, employment and income. Additionally, to capture individuals' cultural background, we include dummies for their religion, ethnicity and the language family associated with their home language. We also include controls for country of residence and WVS wave. These variables are described in Table A1 in the Appendix.

The resulting sample consist of 109,116 individuals in 73 countries, speaking 50 languages, as reported in our baseline regression. 26% of individuals in the sample believe that wife beating can be justifiable. 70% of individuals speak a language with a sex-based gender system.

#### 5 Results

#### 5.1 Baseline results

Our analysis starts with the estimation of equation 2. The marginal effects derived from the model are reported in Table 2.<sup>6</sup> We estimate a parsimonious specification first, limiting the controls to country, wave and language family dummies. The use of country fixed effects

<sup>&</sup>lt;sup>5</sup>We discuss this issue in greater detail in the results section, see Table A5.

<sup>&</sup>lt;sup>6</sup>Since probit as a non-linear model could face an incidental parameters problem with the expansive set of fixed effects in the specification, we also run the analysis using a linear probability model, reporting the results in the Appendix - see Table A2.

controls for a variety of country level factors that might influence individuals' beliefs about IPV, including a country's economic and educational systems, legal and political institutions, as well as national social norms around gender and violence. As argued by Roberts et al. (2015), the inclusion of language family fixed effects helps to control for omitted variables related to the geographical and historical relatedness of languages. The results, presented in column (1), point to a positive and significant association between speaking a language with a sex-based gender system and believing that IPV can be justifiable.

In columns (2) and (3) we augment the specification with additional controls, including individuals' gender, age cohort, marital and parental status, educational attainment, income, and a set of dummy variables that reflect adherence to major world religions. Controlling for religious adherence is particularly important given the strong association between particular religions and languages, such as Catholicism and Spanish and Islam and Arabic, as well as evidence that religion plays an important role in determining attitudes and beliefs related to gender (Davis, 2021). As seen in column 3, the relationship between gendered language and the justifiability of IPV is robust to the inclusion of these variables.

Column (4) presents the most extensive specification where ethnicity dummies are additionally included. Like the inclusion of religious affiliation, the inclusion of ethnic fixed effects is intended to control for ethnicity-level cultural norms that may influence beliefs about IPV.<sup>7</sup> Across all these specifications, we estimate a positive statistically significant marginal effect on SEX-BASED GENDER.<sup>8</sup>

Estimation results from the most extensive model suggest that speaking a gendered language increases the likelihood that an individual believes that IPV is justifiable by 7.5 percentage points. By comparison, the estimated effect of speaking a gendered language on the belief that IPV is justifiable is 27% larger than the difference in beliefs associated with having a college vs. primary education. Alternately, the effect is nearly as large as the impact

<sup>&</sup>lt;sup>7</sup>Admittedly it is challenging to control for all relevant markers of culture. While WVS contains subjective measures of gender attitudes, they are likely to be endogenous to gendered language and their inclusion in the model would therefore lead to the problem of bad controls (Angrist & Pischke, 2009). Our list of proxies for culture is therefore deliberately limited to arguably exogenous measures. In robustness checks reported in the Appendix Table A3 we additionally control for sub-national region dummies available in the WVS, to account for local culture. Our results remain robust.

<sup>&</sup>lt;sup>8</sup>The standard errors in all our models are clustered at the country-level. In Table A4 we show that our results remain robust to alternative approaches to clustering the standard error at the level of language family, country and language family, language, country and language.

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: WIFE BEATING JUSTIFIABLE					
	All	All	All	All	Females	Males
Sex-based gender	0.045**	0.051**	0.066**	0.075***	0.089**	0.056*
	(0.022)	(0.022)	(0.031)	(0.024)	(0.037)	(0.029)
Male		0.078***	0.078***	0.081***		
		(0.006)	(0.006)	(0.007)		
Married		-0.015***	-0.018***	-0.016***	-0.014***	-0.023***
		(0.005)	(0.005)	(0.005)	(0.005)	(0.008)
NO CHILDREN		-0.010	-0.006	-0.004	0.012*	-0.019***
		(0.007)	(0.006)	(0.005)	(0.007)	(0.007)
Primary			0.059***	0.057***	0.055***	0.057***
			(0.010)	(0.010)	(0.011)	(0.011)
Secondary			0.022***	0.023***	0.019**	0.028***
			(0.008)	(0.007)	(0.009)	(0.008)
Employed			-0.001	-0.002	-0.003	0.000
			(0.005)	(0.005)	(0.005)	(0.007)
Age cohort dummies	No	Yes	Yes	Yes	Yes	Yes
Income group dummies	No	No	Yes	Yes	Yes	Yes
Religious denomination dummies	No	No	Yes	Yes	Yes	Yes
Ethnicity dummies	No	No	No	Yes	Yes	Yes
Language family dummies	Yes	Yes	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes
Wave dummies	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R <sup>2</sup>	0.091	0.098	0.096	0.110	0.114	0.103
Mean of dependent	0.259	0.259	0.257	0.262	0.227	0.302
variable						
N	138,932	138,517	124,312	109,116	57,245	51,663

Table 2: Baseline regressions — probit marginal effects

Note.— Robust standard errors clustered by country are in parentheses. \*Denotes significance at 10 percent; \*\*at 5 percent; \*\*\*at 1 percent levels.

of gender itself: male respondents are 8.1 percentage points more likely to report that IPV is justifiable.

In the final two columns of Table 2, we present marginal effect estimates of the specification used in column 4 for subsamples of women and men. In both sets of results, we confirm a positive significant relationship between SEX-BASED GENDER and WIFE BEATING JUSTIFIABLE. Estimation results suggest marginal effects of 8.9 percentage points for women and 5.6 percentage points for men, so that the effect is both larger and more precisely estimated for women. This result is particularly noteworthy given the disproportionate role women play in childrearing and in the transmission of intergenerational values, including those related to gender roles.

## 5.2 Robustness checks

Next, we assess the robustness of our results to the estimation sample employed. The regions of Southern Asia (35%) and Sub-Saharan Africa (33%) have the highest prevalence rates of lifetime IPV after Melanesia, Micronesia and Polynesia that are not in our sample (World Health Organization, 2018). To check whether our results are sensitive to the inclusion of these regions in the sample, we exclude them sequentially from the analyses reported in columns (1) and (2) of Table 3. The positive relationship between SEX-BASED GENDER and WIFE BEATING JUSTIFIABLE persists in both sets of results. Thus, our results do not support the contention that the relationship between gendered language and beliefs about IPV are driven primarily by these regions.

A second concern is that our results may disproportionately reflect the relationship between IPV beliefs and gendered language for a handful of global languages that are overrepresented in our sample, Arabic, Spanish and English. To see if this is so, we sequentially drop these languages from our analysis. As the results reported in columns (3)–(5) show, this does not substantially alter the nature of our results either.

Some existing analyses concerned with the identification of causal effects of linguistic structures on social and economic outcomes focus on the behaviour of immigrants (Gay et al., 2018; Galor et al., 2020). Inspired by the epidemiological approach that exploits the variation across immigrants for the identification of cultural traits on behaviour (Giuliano, 2007; Fernández & Fogli, 2009), these studies extend the approach further to disentangle the impact of linguistic traits from the effect of other ancestral characteristics.<sup>9</sup> In alignment with these studies, we restrict the sample to children of immigrants in the analysis reported in column (6) of Table 3. In spite of the significant drop in the sample size, we estimate a statistically significant positive marginal effect on SEX-BASED GENDER.<sup>10</sup>

Next, we explore whether the variation in the relevant language structures stems from a specific set of countries. Appendix Table A5 shows that the majority of countries in the

<sup>&</sup>lt;sup>9</sup> However, as Beblo et al. (2020) note, individuals who select into migration are more likely to reject the norms of their country of origin and may transmit their traits to their children; thus, the existing estimates of gendered language on migrant behaviour may be biased.

<sup>&</sup>lt;sup>10</sup>We are unfortunately not able to control for the birthplace of parents in this analysis because of a lack of corresponding information in the WVS.

	(1)	(2)	(3)	(4)	(5)	(6)
		Depende	ent variable:	WIFE BEATING	G JUSTIFIABLE	
	No Africa	No Asia	No Arabic	No Spanish	No English	Immigrants
SEX-BASED GENDER	0.077***	0.107***	0.075***	0.081***	0.088***	0.114***
	(0.026)	(0.040)	(0.024)	(0.025)	(0.029)	(0.027)
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Language family dummies	Yes	Yes	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes
Wave dummies	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R <sup>2</sup>	0.105	0.110	0.110	0.117	0.093	0.120
Mean of dependent	0.250	0.239	0.262	0.284	0.281	0.245
variable						
Ν	99,355	82,157	109,116	85,475	93,004	7,384

Table 3: Robustness to sample — probit marginal effects

Note.— Robust standard errors clustered by country are in parentheses. \*Denotes significance at 10 percent; \*\*at 5 percent; \*\*\*at 1 percent levels.

sample exhibit in fact little variation in gendered language, since most countries have a single dominant language (a similar observation is made in the study of language future tense and future-oriented behaviours by Chen (2013)). In fact it is only in a very small number of countries where we observe significant within-county language differences. Yet these countries likely provide the majority of identification in our within-country regressions. We run regressions in sub-samples of countries where the smallest language share is over 1% vs. those where it is 1% or below. Reassuringly, in both sub-samples we retain the significant positive coefficients on SEX-BASED GENDER.

Does gendered language predict the gendered nature of violence, as part of our conceptual framework suggests? The focus of our analysis hitherto has been the beliefs legitimizing physical IPV, i.e. wife beating. Our baseline results, reiterated in column (1) of Table 4, confirm that gendered language is correlated with beliefs legitimising physical IPV. However, gendered language might have implications for beliefs about other forms of gendered violence too, e.g. emotional violence and coercive control. While our dataset does not contain precise measures of these forms of violence, it has information on individuals' views on whether a wife must obey her husband, a social norm that might serve to justify IPV if it is violated. We distinguish between individuals who strongly disagree/disagree with the

(1)	(2)	(3)	(4)	(5)
		Dependent variab	le:	
WIFE BEATING	WIFE MUST OBEY	VIOLENCE AGAINST	CHILD BEATING	POLITICAL VIOLENCE
JUSTIFIABLE	HUSBAND	PEOPLE JUSTIFIABLE	JUSTIFIABLE	JUSTIFIABLE
0.075***	0.030*	0.093	-0.011	0.061
(0.024)	(0.017)	(0.089)	(0.059)	(0.106)
Yes	Yes	Yes	Yes	Yes
Yes	Yes	Yes	Yes	Yes
Yes	Yes	Yes	Yes	Yes
Yes	Yes	Yes	Yes	Yes
0.110	0.188	0.082	0.138	0.138
0.262	0.905	0.309	0.484	0.313
109,116	11,002	83,736	83,806	37,632
	(1) WIFE BEATING JUSTIFIABLE 0.075*** (0.024) Yes Yes Yes Yes 0.110 0.262 109,116	(1)         (2)           WIFE BEATING         WIFE MUST OBEY           JUSTIFIABLE         HUSBAND           0.075***         0.030*           (0.024)         (0.017)           Yes         Yes           Yes         Yes           Yes         Yes           O.110         0.188           0.262         0.905           109,116         11,002	(1)         (2)         (3)           WIFE BEATING         WIFE MUST OBEY         Dependent variable           JUSTIFIABLE         HUSBAND         PEOPLE JUSTIFIABLE           0.075***         0.030*         0.093           (0.024)         (0.017)         (0.089)           Yes         Yes         Yes           O.110         0.188         0.082           0.262         0.905         0.309           109,116         11,002         83,736	(1)         (2)         (3)         (4)           Dependent variable:           WIFE BEATING         WIFE MUST OBEY         VIOLENCE AGAINST         CHILD BEATING           JUSTIFIABLE         HUSBAND         PEOPLE JUSTIFIABLE         JUSTIFIABLE           0.075***         0.030*         0.093         -0.011           (0.024)         (0.017)         (0.089)         (0.059)           Yes         Yes         Yes         Yes           0.110         0.188         0.082         0.138           0.262         0.905         0.309         0.484           109,116         11,002         83,736         83,806

Table 4: Gendered vs. non-gendered violence — probit marginal effects

Note.— Robust standard errors clustered by country are in parentheses. \*Denotes significance at 10 percent; \*\*at 5 percent; \*\*\*at 1 percent levels.

statement vs. those who strongly agree/agree or are agnostic about it.<sup>11</sup> We use this measure as our dependent variable in the regressions reported in the second column of Table 4. The results suggest a positive significant relationship between SEX-BASED GENDER and the probability of agreeing with the statement that a wife must obey her husband.

In light of the cognitive aspect of our conceptual framework, which highlights the psychological role of gender schemata, there is no reason to expect gendered language to be associated with the legitimization of other, non-gendered forms of violence. In the remainder of Table 4, we focus on the justifiability of a range of non-gendered violence forms as outcomes—violence against other people, against children, and for political motives. The results show that SEX-BASED GENDER is statistically unrelated to acceptance of violence against other people (column (3)), child beating (column (4)), and politically-motivated violence (column (5)). Thus gendered language is associated with gendered violence specifically, and not with non-gendered forms of violence.<sup>12</sup>

In additional placebo tests, reported in the Appendix Table A6, we ascertain that we are indeed tracking a relationship associated with the presence of sex-based gender systems

<sup>&</sup>lt;sup>11</sup> This information, however, is only available for the 1999-2004 wave of the survey and for a small group of countries with conservative norms (Algeria, Bangladesh, Indonesia, Iran, Iraq, Jordan, Nigeria, Pakistan, Saudi Arabia, Turkey and Egypt), which explains the high prevalence of views concurring with the statement - see the second last row of Table 4.

<sup>&</sup>lt;sup>12</sup>While the estimated coefficients are not statistically different from zero, the signs and the sizes of the coefficients in columns (3) and (5) of Table 4 are not too dissimilar to those in the baseline specification reported in column (1) of the Table. We cannot therefore firmly conclude that speaking a gendered language is entirely unrelated to justifying other forms of violence.

as opposed to other correlated linguistic or cultural features, by re-running the baseline regression replacing SEX-BASED GENDER with other linguistic features conceptually unrelated with our dependent variable: (i) the presence of periphrastic *future tense* in a language (for engagements with this grammatical feature in other social contexts see Galor et al. (2020); Mavisakalyan et al. (2022)); (ii) whether a language permits speakers to drop a personal pronoun when it is used as the subject of a sentence (*pronoun drop*) and (iii) the presence of informal vs. formal *politeness* differentiation in second person pronouns (see e.g., Tabellini (2008); Davis & Abdurazokzoda (2016) for engagements with these two features in other contexts). The estimated marginal effects on these variables are statistical insignificant (columns 1– 3). We also include SEX-BASED GENDER, FUTURE TENSE, PRONOUN DROP and POLITENESS jointly in a regression presented in column (4). The estimated marginal effect on SEX-BASED GENDER maintains its significance, while those on other grammatical features are insignificant.

Our definition of gendered language is based on the binary distinction between sex-based and non-sex-based gender systems characterised in Corbett (2013b). Previous studies by Hicks et al. (2015) and Gay et al. (2018) have proposed measures of gender intensity in languages utilising the following additional features of grammatical gender in languages: number of genders (NG); gender assignment (GA) and gendered pronouns (GP) .<sup>13</sup> Gay et al. (2018) define the linguistic gender intensity as SEX-BASED\*(GP+GA+NG) and Hicks et al. (2015) define it as SEX-BASED+GP+GA+NG where SEX-BASED is the baseline measure of gendered language used in this study based on the binary distinction between sex-based and non-sex-based gender systems in a language. The intensity measures range from 0-3 and are utilised in the analyses presented in the last two columns of the Appendix Table A6. Utilising these measures of gendered language and finding wife beating justifiable. However, these results need to be approached with caution given that they are not introduced and characterized by

<sup>&</sup>lt;sup>13</sup>NG takes the value of 1 for languages with exactly two genders and a 0 otherwise based on the intuitive premise that "in languages with more than two genders, the presence of additional genders decreases the frequency with which speakers must employ the masculine and feminine gender" (Hicks et al., 2015, p. 24). GA takes the value of 1 for languages whose gender assignment system is both semantic and formal and 0 otherwise. GP equals 1 for languages with gender distinction in third-person pronouns and in the first and/or the second person and 0 if the language does not distinguish gender in pronouns or does so only in the third-person. For detailed motivation and definition of these variables see Hicks et al. (2015) and Gay et al. (2018).

linguists, and therefore our preferred specifications are limited to those using the SEX-BASED measure.

## 5.3 Why gendered language matters: culture vs. cognition

In this section, we provide suggestive insights on why gendered language matters for beliefs about the justifiability of wife beating. As noted in the introduction, speaking a gendered language may be correlated with the intensity of gender roles for two reasons. First, the grammatical structure of a language may directly affect a speaker's cognition and, thus, play a causal role in belief formation and retention. Second, if languages evolve to complement existing cultural values and norms, as argued by Galor et al. (2018), then speaking a gendered language may indicate that gender roles are more significant in an individual's inherited cultural values. In this case, gendered language does not necessarily play a causal role in the formation of an individual's beliefs, but rather indicates that the individual inherited a relatively gender-biased set of cultural values.

Through the introduction of an extensive list of controls which capture individuals' cultural background in the baseline model, we have made an attempt to isolate the influence of language from the influence of correlated cultural factors, and to throw light on the relative significance of cultural vs. cognitive channels of influence. However, comprehensively controlling for culture is a challenging task; at least some of the estimated effect of gendered language is likely to be attributable to culture.

To shed more light on the relative roles of culture and cognition in the relationship between sex-based language and beliefs about the justifiability of wife beating, we utilise an additional approach. We incorporate information on grammatical structure of second languages spoken by individuals into the analysis. In addition to the language used at home, the WVS records information on the language in which the interview is conducted. We exploit the presence of the two languages to provide insight into the relative contributions of the cultural and cognitive channels of influence. In this framework, we assume that the language spoken at home, which like cultural values, is passed down from an individual's parents, is potentially correlated with an individual's inherited cultural beliefs. In contrast, the interview language, which the individual is actively speaking and thinking in at the time of the interview, is more closely associated with the cognitive channel of influence. Earlier studies in social psychology have documented differences in attitudes expressed by subjects attributable to the languages used to elicit these attitudes (Danziger & Ward, 2010; Ogunnaike et al., 2010). Our approach imitates the recent experimental approaches by Ayres et al. (2023) and Pérez & Tavits (2017, 2019). They prompt bilingual speakers, fluent in two languages that vary in their grammatical structure, to make a decision or express an opinion in order to identify the causal effect of language on behavior.

We start by looking at whether and how the gender-intensity of individuals' language repertoires might shape their attitudes on the justifiability of wife beating. To that end, we exploit the information on languages spoken at home and those used to conduct the interviews in to distinguish between individuals who do not speak a gendered language at all, those who speak only one gendered language (but also at least one non-gendered language), and those who speak exclusively gendered languages (omitted). The results reported in column (1) of Table 5 suggest that relative to individuals who only speak a gendered language, those who speak no or a single gendered language have a lower probability of holding an IPV-legitimising belief.

We expand on this estimation further, by introducing a distinction between four groups of individuals based on whether their home and interview languages are gendered. As we see in column (2) of Table 5, moving from a context where all the languages spoken by an individual are gendered to one where only the interview language is gendered is associated with 6.3 percentage points decrease in the probability of reporting an IPV-legitimising belief. However, the estimated marginal effect on ONLY HOME LANGUAGE GENDERED, while negative, is statistically insignificant in this estimation.

In columns (3)–(4), we present results for regressions intended to mimic the experimental design commonly used in psychology. Given the dominant role of home language structure in belief formation, we check whether for individuals whose home language is gendered, using a non-gendered interview language de-activates the beliefs on the justifiability of wife beating. The results estimated in the sub-sample of individuals whose home language has sex-based gender system are presented in column (3) of Table 5. The estimated marginal effect on ONLY HOME LANGUAGE GENDERED is significant and negative, which suggests that eliciting the IPV justifiability beliefs in a non-gendered language among individuals

speaking a gendered language at home might potentially de-activate these beliefs. Note that the p-value for the estimated marginal effect, p = 0.053, narrowly missed the conventional threshold for statistical significance.

In column (4) of Table 5, we check whether for speakers of non-gendered home languages, using a gendered interview language increases acceptance of IPV. Thus, we restrict the estimation sample to speakers of non-gendered languages to obtain the results reported in column (4). We estimate a positive marginal effect on ONLY INTERVIEW LANGUAGE GENDERED, however it is statistically insignificant. Thus, we fail to find evidence of significant cognitive effects of using a gendered interview language on beliefs about wife beating for individuals who speak a non-gendered language at home.

It should be noted that unlike the experimental approaches, the selection into the language of the interview is likely not random. In the case of migrants, it might be, for example, that individuals who take the interview in a country's statutory language are already more socialized to cosmopolitan values, and as such, regard gendered violence as less acceptable, whereas those who still do not feel confident enough in the language of the host country are also less integrated in the local society, and therefore, might have stronger attachments to the values of their country of origin. To address this issue to an extent, in the final two columns of Table 5 we repeat the analysis in columns (3)–(4) in a subsample that excludes children of immigrant parents. The results in this smaller sample mirror those reported in columns (3)–(4).

In summary, our investigation into the role of second languages finds some evidence that the gender structure of the interview language influences beliefs about gendered violence. In particular, for individuals who speak a gendered language at home, the use of genderless interview language may de-activate attitudes supporting such violence. This is consistent with a cognitive effect of language on attitude formation. However, our results also highlight the dominant role of home language gender in shaping the attitudes on gendered violence. For individuals whose home language lacks a sex-based gender system, being interviewed in a gendered language does not seem to be relevant. This result is consistent with complementarity between the cultural and cognitive effects of language.

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: WIFE BEATING JUSTIFIABLE					
	All	All	Sex-based	Sex-based	Sex-based	Sex-based
			GENDER =1	gender =0	GENDER =1 &	GENDER =0 &
					MIGRANT =0	MIGRANT =0
One gendered	-0.048*					
LANGUAGE	(0.026)					
NO GENDERED	-0.074*	-0.088***				
LANGUAGE	(0.030)	(0.024)				
Only home		-0.038	-0.074*		-0.078**	
LANGUAGE GENDERED		(0.038)	(0.038)		(0.035)	
ONLY INTERVIEW		-0.063**		0.021		0.025
LANGUAGE GENDERED		(0.032)		(0.031)		(0.033)
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Language family dummies	Yes	Yes	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes
Wave dummies	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R <sup>2</sup>	0.108	0.108	0.109	0.092	0.109	0.109
Mean of dependent	0.258	0.258	0.231	0.324	0.243	0.329
variable						
Ν	102,587	102,587	73,281	29,231	63,292	21,423

## Table 5: Tests with multilinguals — probit marginal effects

Note.— Robust standard errors clustered by country are in parentheses. \*Denotes significance at 10 percent; \*\*at 5 percent; \*\*\*at 1 percent levels.

Within the limits of a quasi-experimental setup, we regard this as an informative exercise. However, we are not able to convincingly address the issue of selection into the interview language. It is possible that less conservative individuals are more willing to learn a second language but also less likely to believe that wife beating is justifiable. In such cases, the interpretation of the results as pointing to a cognitive rather than cultural effect might be too strong, and should be handled with caution. In addition, those interviewed in a country's statutory language, often the language of a Western colonizer, may feel less comfortable expressing opinions that conflict with cosmopolitan values. It should be noted, however, that many colonial languages are gendered, including French and Spanish, so speaking a country's statutory language maps imperfectly onto being interviewed in a non-gendered language.

## 6 Conclusion

This paper provides empirical evidence linking gendered language and gendered violence. We characterize a language as gendered if it has a gender system whose semantic core is based on biological sex. According to a cognitive mechanism, speaking a gendered language may lead to a higher activation of gender schemata in the minds of speakers, increasing the significance of gender distinctions and existing gender norms, potentially leading to the legitimization of gender inequality, including gendered violence. Alternatively, according to a cultural mechanism, gendered language serves as a marker for underlying gendered social norms.

Using cross country data, we identify a large, statistically significant relationship between gendered language and the incidence of IPV. Controlling for a set of common economic, political, religious and institutional variables, we find that having a gendered dominant language is associated with an 8.5 percentage point increase in the share of women who experience IPV.

Motivated by this association, we use individual level data from the WVS to explore the link between speaking a gendered language and the belief in the justifiability of wife beating. In our preferred specification, we find that speaking a gendered language is associated with a 7.5 percentage point increase in the likelihood that an individual believes that wife beating is justifiable. This specification controls for a wide variety of individual level socioeconomic characteristics as well as country, religion, language family and ethnicity fixed effects. Our primary finding is also robust to the use of samples that exclude high-IPV regions and global languages. In line with the described cognitive mechanism, which highlights the role of gender schemata, we fail to find a significant relationship between gendered language and beliefs about other, non-gendered forms of violence.

Finally, we exploit evidence on the beliefs of multilingual individuals to further distinguish between the cultural and the cognitive channel of influence. In particular, we find that for respondents who speak a gendered language at home, being interviewed in a nongendered language significantly reduces the acceptance of wife beating, which provides some evidence in favour of the cognitive channel of influence, and as supporting the potential efficacy of language-based policy interventions to reduce IPV.

#### References

- Aizer, Anna. 2010. The gender wage gap and domestic violence. American economic review, 100(4), 1847–59.
- Aizer, Anna. 2011. Poverty, violence, and health: the impact of domestic violence during pregnancy on newborn health. Journal of human resources, 46(3), 518–538.
- Alesina, Alberto, Devleeschauwer, Arnaud, Easterly, William, Kurlat, Sergio, Wacziarg, Romain. 2003. Fractionalization. Journal of economic growth, 8(2), 155–194.
- Alesina, Alberto, Brioschi, Benedetta, La Ferrara, Eliana. 2021. Violence against women: A cross-cultural analysis for Africa. Economica, 88(349), 70–104.
- Amaral, Sofia. 2017. Do improved property rights decrease violence against women in India? ISER Working Paper Series. URL: https://www.econstor.eu/bitstream/10419/ 200362/1/1002513332.pdf.
- Amaral, Sofia, Bhalotra, Sonia R, Prakash, Nishith, et al. 2021. Gender, crime and punishment: Evidence from women police stations in India. Institute of Labor Economics (IZA) Working Paper. URL: http://ftp.iza.org/dp14250.pdf.
- Anderberg, Dan, Moroni, Gloria. 2020. Exposure to intimate partner violence and children's dynamic skill accumulation: Evidence from a UK longitudinal study. Oxford review of economic policy, 36(4), 783–815.
- Angrist, Joshua D, Pischke, Jörn-Steffen. 2009. Mostly harmless econometrics: An empiricist's companion. Princeton university press.
- Arenas-Arroyo, Esther, Fernandez-Kranz, Daniel, Nollenberger, Natalia. 2021. Intimate partner violence under forced cohabitation and economic stress: Evidence from the COVID-19 pandemic. Journal of public economics, 194, 104350.
- Atkinson, Maxine P, Greenstein, Theodore N., Lang, Molly M. 2005. For women, breadwinning can be dangerous: Gendered resource theory and wife abuse. Journal of marriage and family, 67, 1137–1148.
- Ayres, Ian, Katz, Tamar Kricheil, Regev, Tali. 2023. Languages and future-oriented economic behavior—experimental evidence for causal effects. Proceedings of the national academy of sciences, 120(7), e2208871120.

- Baranowski, Maciej. 2002. Current usage of the epicene pronoun in written english. Journal of sociolinguistics, 6(3), 378–397.
- Bargain, Olivier, Boutin, Delphine, Champeaux, Hugues. 2019. Women's political participation and intra-household empowerment: Evidence from the Egyptian Arab Spring. Journal of development economics, 141, 102379.
- Barro, Robert. 2003. Religion adherence data. accessed at: https://scholar.harvard.edu/barro/publications/religion-adherence-data.
- Beblo, Miriam, Görges, Luise, Markowsky, Eva. 2020. Gender matters in language and economic behaviour. Labour economics, 101850.
- Beleche, Trinidad. 2019. Domestic violence laws and suicide in Mexico. Review of economics of the household, 17(1), 229–248.
- Beltrán Tapia, Francisco J, Gallego-Martínez, Domingo. 2020. What explains the missing girls in nineteenth-century Spain? The economic history review, 73(1), 59–77.
- Berniell, Inés, Facchini, Gabriel. 2021. COVID-19 lockdown and domestic violence: Evidence from internet-search behavior in 11 countries. European economic review, 103775.
- Bharati, Tushar, Mavisakalyan, Astghik, Vu, Loan. 2024. Intimate partner abuse and child health. Tech. rept. GLO Discussion Paper.
- Borker, Girija. 2017. Safety first: Perceived risk of street harassment and educational choices of women. Job Market Paper, Department of Economics, Brown University. URL: https: //girijaborker.files.wordpress.com/2020/11/borker\_2020\_safetyfirst-1.pdf.
- Boroditsky, Lera. 2003. Linguistic relativity. Encyclopedia of cognitive science.
- Boroditsky, Lera, Schmidt, Lauren A, Phillips, Webb. 2003. Sex, syntax, and semantics. Pages 61–79 of: Gentner, Dedre, Goldin-Meadow, Susan (eds), Language in mind: Advances in the study of language and thought. Cambridge: MIT press.
- Campbell, Jacquelyn C. 2002. Health consequences of intimate partner violence. The lancet, 359(9314), 1331–1336.
- Casasanto, Daniel. 2015. Linguistic relativity. Pages 158–174 of: Riemer, N. (ed), Routledge handbook of semantics. New York: Routledge.
- Chen, M. Keith. 2013. The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. The american economic review, 103(2),

690-731.

- Comrie, Bernard, Dryer, Matthew S., Gil, David, Haspelmath, Martin. 2013. Introduction (v2020.3). In: Dryer, Matthew S., Haspelmath, Martin (eds), The world atlas of language structures online. Zenodo.
- Cools, Sara, Flatø, Martin, Kotsadam, Andreas. 2020. Rainfall shocks and intimate partner violence in sub-saharan africa. Journal of peace research, 57(3), 377–390.
- Corbett, G.G. 1991. Gender. Cambridge Textbooks in Linguistics. Cambridge University Press.
- Corbett, Greville G. 2013a. Number of genders. In: Dryer, Matthew S., Haspelmath, Martin (eds), The world atlas of language structures online. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Corbett, Greville G. 2013b. Sex-based and non-sex-based gender systems (v2020.3). In: Dryer, Matthew S., Haspelmath, Martin (eds), The world atlas of language structures online. Zenodo.
- Corbett, Greville G. 2013c. Systems of gender assignment. In: Dryer, Matthew S., Haspelmath, Martin (eds), The world atlas of language structures online. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Currie, Janet, Mueller-Smith, Michael, Rossin-Slater, Maya. 2018. Violence while in utero: The impact of assaults during pregnancy on birth outcomes. The review of economics and statistics, 1–46.
- Dahl, Osten. 2013. Stuck in the futureless zone. Diversity linguistics comment posted 03/09/2013.
- Danziger, Shai, Ward, Robert. 2010. Language changes implicit associations between ethnic groups and evaluation in bilinguals. Psychological science, 21(6), 799.
- Davis, Lewis. 2021. Culture, development, and the divergence of women's empowerment. Manuscript, Union College.
- Davis, Lewis, Reynolds, Megan. 2018. Gendered language and the educational gender gap. Economics letters, 168, 46–48.
- Davis, Lewis S, Abdurazokzoda, Farangis. 2016. Language, culture and institutions: Evidence from a new linguistic dataset. Journal of comparative economics, 44(3), 541–561.

- Devries, Karen M, Mak, Joelle Y, Bacchus, Loraine J, Child, Jennifer C, Falder, Gail, Petzold, Max, Astbury, Jill, Watts, Charlotte H. 2013. Intimate partner violence and incident depressive symptoms and suicide attempts: A systematic review of longitudinal studies. Plos medicine, 10(5), e1001439.
- Dryer, Matthew S., Haspelmath, Martin (eds). 2013. Wals online. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Duvvury, Nata, Callan, Aoife, Carney, Patrick, Raghavendra, Srinivas. 2013. Intimate partner violence: Economic costs and implications for growth and development. World Bank, Washington, DC. URL: https://documents.worldbank.org/curated/en/ 412091468337843649/pdf/825320WP0Intim00Box379862B00PUBLIC0.pdf.
- Eberhard, David M., Gary F. Simons, Fennig, Charles D. 2024. Ethnologue: Languages of the world. Dallas, Texas: SIL International. Online version.
- Fernández, Raquel, Fogli, Alessandra. 2009. Culture: An empirical investigation of beliefs, work, and fertility. American economic journal: Macroeconomics, 1(1), 146–77.
- Fiske, Susan, Taylor, Shelley. 1991. Social cognition. New York: McGraw- Hill Book Company.
- Freedom House. 2022. Freedom in the world 2022. accessed at: https://freedomhouse.org/report/freedom-world.
- Galor, Oded, Özak, Ömer, Sarid, Assaf. 2018. Geographical roots of the coevolution of cultural and linguistic traits. Nber working paper.
- Galor, Oded, Özak, Ömer, Sarid, Assaf. 2020. Linguistic traits and human capital formation. Aea papers and proceedings.
- García-Ramos, Aixa. 2021. Divorce laws and intimate partner violence: Evidence from Mexico. Journal of development economics, 150, 102623.
- Gay, Victor, Hicks, Daniel L, Santacreu-Vasut, Estefania, Shoham, Amir. 2018. Decomposing culture: an analysis of gender, language, and labor supply in the household. Review of economics of the household, 16(4), 879–909.
- Georgetown Institute for Women, Peace and Security and Peace Research Institute Oslo. 2019. Women, peace and security index 2019/20: Tracking sustainable peace through inclusion, justice, and security for women. Washington, DC: GIWPS and PRIO.

- Ginsburgh, Victor, Weber, Shlomo. 2020. The economics of language. Journal of economic literature, 58(2), 348–404.
- Giuliano, Paola. 2007. Living arrangements in western europe: Does cultural origin matter? Journal of the european economic association, 5(5), 927–952.
- Gleitman, Lila, Papafragou, Anna. 2013. Relations between language and thought. Pages 504–523 of: Reisberg, D. (ed), Handbook of cognitive psychology. Oxford University Press.
- Gumperz, John J, Levinson, Stephen. 1996. Rethinking linguistic relativity. Studies in the Social and Cultural Foundations of Language, vol. 17. Cambridge: Cambridge University Press.
- Heath, Rachel. 2014. Women's access to labor market opportunities, control of household resources, and domestic violence: Evidence from Bangladesh. World development, 57, 32–46.
- Heise, Lori L, Kotsadam, Andreas. 2015. Cross-national and multilevel correlates of partner violence: an analysis of data from population-based surveys. The lancet global health, 3(6), e332–e340.
- Hicks, Daniel L, Santacreu-Vasut, Estefania, Shoham, Amir. 2015. Does mother tongue make for women's work? linguistics, household labor, and gender identity. Journal of economic behavior & organization, 110, 19–44.
- Hidrobo, Melissa, Fernald, Lia. 2013. Cash transfers and domestic violence. Journal of health economics, 32(1), 304–319.
- Hidrobo, Melissa, Peterman, Amber, Heise, Lori. 2016. The effect of cash, vouchers, and food transfers on intimate partner violence: Evidence from a randomized experiment in Northern Ecuador. American economic journal: Applied economics, 8(3), 284–303.
- Jacoby, Hanan G, Mansuri, Ghazala. 2010. Watta satta: Bride exchange and women's welfare in rural Pakistan. American economic review, 100(4), 1804–25.
- Jakiela, Pamela, Ozier, Owen. 2018. Gendered language. World bank policy research working paper.
- Jayachandran, Seema. 2015. The roots of gender inequality in developing countries. Annual reviews of economics, 7(1), 63–88.

- Kay, Paul, Regier, Terry. 2006. Language, thought and color: Recent developments. Trends in cognitive sciences, 10(2), 51–54.
- Khalil, Umair, Mookerjee, Sulagna. 2019. Patrilocal residence and women's social status: Evidence from South Asia. Economic development and cultural change, 67(2), 401–438.
- Kotsadam, Andreas, Østby, Gudrun, Rustad, Siri Aas. 2017. Structural change and wife abuse: A disaggregated study of mineral mining and domestic violence in sub-Saharan Africa, 1999–2013. Political geography, 56, 53–65.
- La Mattina, Giulia. 2017. Civil conflict, domestic violence and intra-household bargaining in post-genocide Rwanda. Journal of development economics, 124, 168–198.
- La Porta, Rafael, Lopez-de Silanes, Florencio, Shleifer, Andrei. 2008. The economic consequences of legal origins. Journal of economic literature, 46(2), 285–332.
- Levinson, Stephen C, Wilkins, David P. 2006. Grammars of space: Explorations in cognitive diversity. Vol. 6. Cambridge: Cambridge University Press.
- Leyaro, V, Selaya, P, Trifkovic, N, et al. 2017. Fishermen's wives: On the cultural origins of violence against women. UNU WIDER Working Paper, 205. URL: https://www.wider. unu.edu/sites/default/files/Publications/Working-paper/PDF/wp2017-205.pdf.
- Lucy, John A. 1996. Grammatical categories and cognition: A case study of the linguistic relativity hypothesis. Cambridge University Press.
- Lucy, John A. 1997. Linguistic relativity. Annual review of anthropology, 26, 291–312.
- Martin, Carol Lynn, Halverson Jr, Charles F. 1981. A schematic processing model of sex typing and stereotyping in children. Child development, 1119–1134.
- Martinez, M. Anais, Osornio, Alisha, Halim, May Ling D., Zosuls, Kristina M. 2020. Gender: Awareness, identity, and stereotyping. Pages 1–12 of: Benson, Janette B. (ed), Encyclopedia of infant and early childhood development (second edition). Oxford: Elsevier.
- Mavisakalyan, Astghik. 2015. Gender in language and gender in employment. Oxford development studies, 43(4), 403–424.
- Mavisakalyan, Astghik, Minasyan, Anna. 2023. The role of conflict in sex discrimination: The case of missing girls. Economic development and cultural change.
- Mavisakalyan, Astghik, Rammohan, Anu. 2021. Female autonomy in household decisionmaking and intimate partner violence: Evidence from pakistan. Review of economics of

the household, 19(1), 255–280.

- Mavisakalyan, Astghik, Weber, Clas. 2017. Linguistic structures and economic outcomes. Journal of economic surveys, 32(3), 916–939.
- Mavisakalyan, Astghik, Tarverdi, Yashar, Weber, Clas. 2022. Heaven can wait: future tense and religiosity. Journal of population economics, 35(3), 833–860.
- McCoubrey, Carmel. 2017. Toppling the grammar patriarchy. The new york times. available online at: https://www.nytimes.com/2017/11/16/opinion/french-sexismgrammar-everybody.html.
- Minsky, Marvin Lee. 1974. A framework for representing knowledge. Chap. 5, pages 111–142 of: Haugeland, John (ed), Mind design: Philosophy, psychology, artificial intelligence.
  Cambridge: MIT Press.
- Ogunnaike, Oludamini, Dunham, Yarrow, Banaji, Mahzarin R. 2010. The language of implicit preferences. Journal of experimental social psychology, 46(6), 999–1003.
- Oh, K. J. 2003. Language, cognitivion, and development: Motion events in english and korean. Ph.D. thesis, University of California, Berkeley.
- Pérez, Efrén O, Tavits, Margit. 2017. Language shapes people's time perspective and support for future-oriented policies. American journal of political science, 61(3), 715–727.
- Pérez, Efrén O, Tavits, Margit. 2019. Language influences public attitudes toward gender equality. The journal of politics, 81(1), 81–93.
- Pinker, Steven. 1994. The language instinct: How the mind creates language. New York: Harper.
- Politi, Daniel. 2020. In argentina, a bid to make language gender neutral gains traction. The new york times. available online at: https://www. nytimes. com/2020/04/15/world/americas/argentina-gender-language. html (accessed november 12, 2021).
- Roberts, Sean, Winters, James. 2013. Linguistic diversity and traffic accidents: Lessons from statistical studies of cultural traits. Plos one, 8(8), 1–13.
- Roberts, Sean G., Winters, James, Chen, Keith. 2015. Future tense and economic decisions: Controlling for cultural evolution. Plos one, 10(7), 1–46.

- Rosch, Eleanor. 1975. Cognitive representations of semantic categories. Journal of experimental psychology: General, 104(3), 192–233.
- Rosch, Eleanor. 1978. Principles of categorization. Pages 27–48 of: Rosch, Eleanor, Lloyd, Barbara B (eds), Cognition and categorization. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Ruble, Diane N, Martin, Carol Lynn, Berenbaum, Sheri A. 2006. Gender development. John Wiley & Sons, Inc.
- Rumelhart, David E. 1980. Schemata: The building blocks of cognition. Pages 33–58 of: Spiro, R., Bruce, B., Brewer, W. (eds), Theoretical issues in reading comprehension. Erlbaum.
- Santacreu-Vasut, E., Shenkar, O., Shoham, A. 2014. Linguistic gender marking and its IB ramifications. Journal of international business studies, 45(9), 1170–1178.
- Santacreu-Vasut, Estefania, Shoham, Amir, Gay, Victor. 2013. Do female/male distinctions in language matter? Evidence from gender political quotas. Applied economics letters, 20(5), 495–498.
- Saul, Jennifer. 2010. Feminist Philosophy of Language. In: Zalta, Edward N. (ed), The Stanford encyclopedia of philosophy, Fall 2010 edn. Metaphysics Research Lab, Stanford University.
- Schank, Roger C., Abelson, Robert P. 1977. Scripts, plans, goals, and understanding: An inquiry into human knowledge structures. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Slobin, Dan I. 2003. Language and thought online: Cognitive consequences of linguistic relativity. Pages 157–192 of: Gentner, D., Goldin-Meadow, S. (eds), Language in mind: Advances in the study of language and thought. Cambridge: MIT press.
- Tabellini, Guido. 2008. Institutions and culture. Journal of the european economic association, 6(2-3), 255–294.
- Tagliabue, Johnl. 2012. Swedish school's begins with big lesson dropping personal The new vork times. available online pronouns. at: https://www.nytimes.com/2012/11/14/world/europe/swedish-school-deemphasizes-gender-lines.html.

- Tajfel, Henri. 1981. Human groups and social categories: Studies in social psychology. Cambridge: Cambridge University Press.
- Tur-Prats, Ana. 2019. Family types and intimate partner violence: A historical perspective. Review of economics and statistics, 101(5), 878–891.
- Turner, John C., Hogg, Michael A., Oakes, Penelope J., Reicher, Stephen D., Wetherell, Margaret S. 1987. Rediscovering the social group: A self-categorization theory. Oxford: Blackwell Publishers Ltd.
- UN Women. 2016. Re-The economic costs of violence against women. UN Deputy mark by Assistant Secretary-General and Executive Direc-URL:https://www.unwomen.org/en/news/stories/2016/9/ Lakshmi Puri tor, speech-by-lakshmi-puri-on-economic-costs-of-violence-against-women.
- United Nations. 2020. The impact of COVID-19 on women. Policy Brief. URL:https://www. unwomen.org/-/media/headquarters/attachments/sections/library/publications/ 2020/policy-brief-the-impact-of-covid-19-on-women-en.pdf?la=en&vs=1406.
- van der Velde, Lucas, Tyrowicz, Joanna, Siwinska, Joanna. 2015. Language and (the estimates of) the gender wage gap. Economics letters, 136, 165 – 170.
- Whorf, Benjamin Lee. 1956. Language, thought, and reality: Selected writings of benjamin lee whorf. MIT Press.
- Winawer, Jonathan, Witthoft, Nathan, Frank, Michael C, Wu, Lisa, Wade, Alex R, Boroditsky, Lera. 2007. Russian blues reveal effects of language on color discrimination. Proceedings of the national academy of sciences, 114(19), 7780–7785.
- World Bank. 2022. World development indicators. accessed at: https://databank.worldbank.org/source/world-development-indicators.
- World Health Organization. 2018. Violence against women prevalence estimates, 2018. World Health Organization Report. URL: https://apps.who.int/iris/bitstream/ handle/10665/341337/9789240022256-eng.pdf?sequence=1.
- Zhang, Yinjunjie, Breunig, Robert V. 2021. Gender norms and domestic abuse: Evidence from Australia. Tax and Transfer Policy Institute Working paper. URL: https: //taxpolicy.crawford.anu.edu.au/sites/default/files/publication/taxstudies\_ crawford\_anu\_edu\_au/2021-04/complete\_breunig\_zhang\_wp\_mar\_2021.pdf.

## Appendix

Variable	Definition of Variable	Mean	S.D.
WIFE BEATING JUSTIFIABLE	0-1 binary variable; equals 1 if respondent agrees that wife beating	0.262	0.440
Sex-based gender	0-1 binary variable; equals 1 respondent's language has sex-based grammatical gender	0.701	0.458
BASELINE CONTROLS:			
MALE	0-1 binary variable; equals 1 if respondent is male	0.474	0.499
Age under 20	0–1 binary variable; equals 1 if respondent is aged under 20	0.050	0.218
Age 20-29	0–1 binary variable; equals 1 if respondent is aged 20–29	0.239	0.426
Age 30-39	0–1 binary variable; equals 1 if respondent is aged 30–39	0.211	0.408
Age 40-49	0–1 binary variable; equals 1 if respondent is aged 40–49	0.182	0.386
Age 50-59	0–1 binary variable; equals 1 if respondent is aged 50–59	0.150	0.357
Age 60-69	0–1 binary variable; equals 1 if respondent is aged 60–69	0.104	0.305
Age 70-79	0–1 binary variable; equals 1 if respondent is aged 70–79	0.050	0.218
Age 80-89	0–1 binary variable; equals 1 if respondent is aged 80–89	0.013	0.113
Age over 90	0–1 binary variable; equals 1 if respondent is aged over 90	0.001	0.029
MARRIED	0-1 binary variable; equals 1 if respondent is married or cohabiting	0.620	0.485
NO CHILDREN	0-1 binary variable; equals 1 if respondent has no children	0.284	0.451
PRIMARY	0-1 binary variable; equals 1 if respondent has primary-level education	0.245	0.430
Secondary	0-1 binary variable; equals 1 if respondent has secondary-level education	0.474	0.499
TERTIARY	0-1 binary variable; equals 1 if respondent has tertiary-level education	0.280	0.449
Employed	0-1 binary variable; equals 1 if respondent is employed	0.569	0.495
INCOME GROUP 1	0-1 binary variable denoting self-assessed income standing	0.143	0.351
INCOME GROUP 2	0-1 binary variable denoting self-assessed income standing	0.253	0.435
INCOME GROUP 3	0-1 binary variable denoting self-assessed income standing	0.381	0.486
INCOME GROUP 4	0-1 binary variable denoting self-assessed income standing	0.182	0.386
INCOME GROUP 5	0-1 binary variable denoting self-assessed income standing	0.040	0.195
CATHOLIC	0-1 binary variable; equals 1 if respondent has Catholic denomination	0.207	0.405
Protestant	0-1 binary variable; equals 1 if respondent has Protestant denomination	0.081	0.272
Orthodox	0-1 binary variable; equals 1 if respondent has Orthodox denomination	0.108	0.311
OTHER CHRISTIAN	0-1 binary variable; equals 1 if respondent has other Christian denomination	0.047	0.211
Jewish	0-1 binary variable; equals 1 if respondent has Jewish denomination	0.010	0.098
MUSLIM	0-1 binary variable; equals 1 if respondent has Muslim denomination	0.189	0.391
Hundu	0-1 binary variable; equals 1 if respondent has Hindu denomination	0.020	0.139
Buddhist	0-1 binary variable; equals 1 if respondent has Buddhist denomination	0.047	0.211
NO DENOMINATION	0-1 binary variable; equals 1 if respondent has no denomination	0.211	0.408
Other	0-1 binary variable; equals 1 if respondent has other or no denomination	0.081	0.272

Note. — N = 109, 116.

	(1)	(2)	(3)	(4)	(5)	(6)
	De	ependent v	ariable: W	IFE BEATIN	G JUSTIFIA	BLE
	All	All	All	All	Females	Males
Sex-based gender	0.053**	0.059***	0.077**	0.084***	0.099***	0.063**
	(0.021)	(0.021)	(0.034)	(0.022)	(0.032)	(0.027)
MALE		0.074***	0.074***	0.075***		
		(0.006)	(0.006)	(0.007)		
Married		-0.014***	-0.017***	-0.015***	-0.013***	-0.020***
		(0.005)	(0.004)	(0.004)	(0.004)	(0.007)
NO CHILDREN		-0.010	-0.006	-0.004	0.011*	-0.017***
		(0.007)	(0.006)	(0.005)	(0.006)	(0.006)
Primary			0.054***	0.052***	0.052***	0.051***
			(0.009)	(0.009)	(0.010)	(0.011)
Secondary			0.020***	0.021***	0.018**	0.025***
			(0.007)	(0.006)	(0.008)	(0.007)
Employed			-0.000	-0.001	-0.004	-0.000
			(0.005)	(0.004)	(0.005)	(0.007)
Age cohort dummies	No	Yes	Yes	Yes	Yes	Yes
Income group dummies	No	No	Yes	Yes	Yes	Yes
Religious denomination dummies	No	No	Yes	Yes	Yes	Yes
Ethnicity dummies	No	No	No	Yes	Yes	Yes
Language family dummies	Yes	Yes	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes
Wave dummies	Yes	Yes	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.102	0.110	0.106	0.126	0.126	0.125
Mean of dependent	0.259	0.259	0.257	0.262	0.226	0.302
variable						
N	138,936	138,521	124,314	109,313	57,495	51,818

Table A2: Baseline regressions — OLS coefficients

Note.— Robust standard errors clustered by country are in parentheses. \*Denotes significance at 10 percent; \*\*at 5 percent; \*\*\*at 1 percent levels.

	(1)	(2)	(3)	
	Dependent variable:			
	WIFE BEATING JUSTIFIAB			
	All	Females	Males	
Sex-based gender	0.066***	0.066*	0.066**	
	(0.020)	(0.034)	(0.031)	
MALE	0.083***			
	(0.007)			
MARRIED	-0.017***	-0.015***	-0.027***	
	(0.005)	(0.005)	(0.008)	
NO CHILDREN	-0.005	0.010	-0.020***	
	(0.005)	(0.006)	(0.007)	
PRIMARY	0.056***	0.054***	0.058***	
	(0.010)	(0.011)	(0.012)	
Secondary	0.021***	0.015*	0.030***	
	(0.007)	(0.008)	(0.008)	
Employed	0.001	-0.001	0.005	
	(0.004)	(0.005)	(0.008)	
Age cohort dummies	Yes	Yes	Yes	
Income group dummies	Yes	Yes	Yes	
Religious denomination dummies	Yes	Yes	Yes	
Ethnicity dummies	Yes	Yes	Yes	
Language family dummies	Yes	Yes	Yes	
Subnational region dummies	Yes	Yes	Yes	
Wave dummies	Yes	Yes	Yes	
Pseudo R <sup>2</sup>	0.145	0.154	0.142	
Mean of dependent	0.261	0.228	0.303	
variable				
N	106,160	54,524	49,994	

Table A3: Robustness to controlling for sub-national regions — probit marginal effects

Note.— Robust standard errors clustered by country are in parentheses. \*Denotes significance at 10 percent; \*\*at 5 percent; \*\*\*at 1 percent levels.

Table A4:	Tests	with	alternative	approa	ches to	clustering	standard	errors —	· probit
marginal o	effects					C C			-

	(1)	(2)	(3)	(4)	(5)		
		Robust standard errors clustered at the level of:					
	Country	LANGUAGE	COUNTRY AND	LANGUAGE	COUNTRY AND		
		FAMILY	LANGUAGE FAMILY		LANGUAGE		
Sex-based gender	0.075***	0.075***	0.075***	0.075***	0.075***		
	(0.024)	(0.023)	(0.027)	(0.021)	(0.025)		
Baseline controls	Yes	Yes	Yes	Yes	Yes		
Language family dummies	Yes	Yes	Yes	Yes	Yes		
Country dummies	Yes	Yes	Yes	Yes	Yes		
Wave dummies	Yes	Yes	Yes	Yes	Yes		
Number of clusters	73	12	138	50	237		
Pseudo <i>R</i> <sup>2</sup>	0.110	0.110	0.110	0.110	0.110		
Mean of dependent	0.262	0.262	0.262	0.262	0.262		
variable							
N	109,116	11,002	83,736	83,806	37,632		

Note.— \*Denotes significance at 10 percent; \*\*at 5 percent; \*\*\*at 1 percent levels.

(2)	(3)	(4)	(5)
	Smallest lang	uage share $> 1$	
$1 < \text{share} \le 2$	$2 < \text{share} \le 5$	$5 < \text{share} \le 10$	share $> 10$
Australia	Azerbaijan	Bolivia	Cyprus
Hong Kong	Armenia	Uzbekistan	Ghana
Iran	Finland	Zambia	Kyrgyzstan

Kazakhstan

Mali

Peru

Zimbabwe

## Table A5: Within-country language differences

New Zealand

Tajikistan

(1)Smallest language share  $\leq 1$ 

Algeria, Andorra, Argentina,

Bulgaria, Belarus, Canada,

Chile, China, Colombia,

Ecuador, Ethiopia, Estonia,

France, Germany, Greece,

Guatemala, Haiti, Hungary,

India, Indonesia, Iraq,

Jordan, Lebanon, Libya,

Macau, Mexico, Moldova, Morocco, Myanmar, Netherlands, Nicaragua, Pakistan, Philippines, Poland, Puerto Rico, Romania, Russia, Serbia, Vietnam, Sweden, Taiwan, Thailand, Trinidad&Tobago, Tunisia, Ukraine, Egypt, United Kingdom, United States, Uruguay, Yemen

COUNTRIES:

	Smallest language share $\leq 1$	Smallest language share >1	
Sex-based gender	0.343*	0.042*	
	(0.185)	0.025	
Baseline controls	Yes	Yes	
Language family dummies	Yes	Yes	
Country dummies	Yes	Yes	
Wave dummies	Yes	Yes	
Pseudo R <sup>2</sup>	0.102	0.124	
Mean of dependent	0.245	0.299	
N	73,745	35,356	

Note.— Robust standard errors clustered by country are in parentheses. \*Denotes significance at 10 percent; \*\*at 5 percent; \*\*\*at 1 percent levels.

Malaysia

Nigeria

Singapore

South Africa

Burkina Faso

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: WIFE BEATING JUSTIFIABLE					
Tense	-0.016			0.004		
	(0.022)			(0.019)		
Pronoun drop		-0.002		-0.003		
		(0.022)		(0.025)		
Politeness		· · · ·	0.002	0.034		
			(0.029)	(0.026)		
SEX-BASED GENDER				0.092***		
				(0.028)		
GENDER INTENSITY 1 (GAY ET AL., 2018)				(010-0)	0.016**	
					(0.007)	
GENDER INTENSITY 2 (HICKS ET AL 2015)					(0.007)	0 017***
						(0.006)
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Language family dummies	Yes	Yes	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes
Wave dummies	Voc	Voc	Vos	Voc	Vos	Voc
Pooudo $P^2$	0 100	0 1 2 4	0.115	0.112	0.114	0.114
Neen of demondent	0.109	0.124	0.115	0.112	0.114	0.114
	0.248	0.252	0.240	0.247	0.249	0.249
variable	110 050		114 (00	00.000	05 1 4 4	
IN	119,358	119,559	114,682	92,922	95,144	95,144

Table A6: Tests with alternative features of languages — probit marginal effects

Note.— Robust standard errors clustered by country are in parentheses. \*Denotes significance at 10 percent; \*\*at 5 percent; \*\*\*at 1 percent levels.