

Gendered Language

Pamela Jakiela and Owen Ozier

Abstract

Languages use different systems for classifying nouns. Gender languages assign many—sometimes all—nouns to distinct sex-based categories, masculine and feminine. We construct a new data set, documenting this property for more than four thousand languages which together account for more than 99 percent of the world's population. At the cross-country level, we find a robust negative relationship between prevalence of gender languages and women's labor force participation. We also show that traditional views of gender roles are more common in countries with more native speakers of gender languages. Our cross-country data also permit a novel permutation test, demonstrating that the patterns we find are robust to statistical correction for correlation in linguistic structure within language families. We also conduct within-country analysis in two regions where indigenous languages vary in terms of their gender structure. In four countries in Sub-Saharan Africa and in India, we show that educational attainment and female labor force participation are lower among those whose native languages use grammatical gender.

Keywords: grammatical gender, language, gender, linguistic determinism, labor force participation, educational attainment, gender gaps

JEL: J16, Z10, Z13

Gendered Language

Pamela Jakiela

Center for Global Development, University of Maryland,
BREAD, and IZA
pjakiela@cgdev.org

Owen Ozier

World Bank Development Research Group, BREAD, and IZA
oozier@worldbank.org

We are grateful to the Gender Innovation Lab at the World Bank for funding, to Laura Kincaide, Yujie Lin, and Kattya Quiroga Velasco for research assistance, and to Arun Advani, Sarah Baird, Michal Bauer, Lori Beaman, Douglas Bernheim, Hoyt Bleakley, Premila and Satish Chand, Sameer Chand, Keith Chen, Julie Chytilová, Michael Clemens, Austin Davis, Giles Dickenson-Jones, Penny Eckert, Alice Evans, David Evans, Marcel Fafchamps, James Fenske, Deon Filmer, Jed Friedman, Julio Garin, Garance Genicot, Xavi Giné, Jess Goldberg, Markus Goldstein, Guy Grossman, Kyaw Hlaw, Karla Hoff, Guido Imbens, Clement Imbert, Anett John, Shareen Joshi, Eeshani Kandpal, Madhulika Khanna, Brent Kreider, Eliana La Ferrara, Ken Leonard, Robyn Meeks, Margreet Luth-Morgan, Jeremy Magruder, Andy Marshall, Matthew Masten, Justin McCrary, David McKenzie, Ted Miguel, Hannes Mueller, Oyebola Okunogbe, Jessica Olney, A.K. Rahim, Uday Raj, Martin Ravallion, Bob Rijkers, Jesse Rothstein, Justin Sandefur, Fré Schreiber, Pieter Serneels, Duncan Thomas, Dominique van de Walle, Tom Wasow, Andrew Zeitlin, and seminar audiences at Bocconi, CERGE-EI (Prague), CSAE (Oxford), Duke, the European University Institute, Georgetown, the Institute for Fiscal Studies, Stanford, UAB (Barcelona), the University of Delaware, the University of Michigan, the University of Warwick, and the World Bank for helpful comments. All errors are our own. The findings, interpretations and conclusions expressed in this paper are entirely those of the authors, and do not necessarily represent the views of the World Bank, its Executive Directors, or the governments of the countries they represent.

Read the online appendix: <http://www.pamjakiela.com/JakielaOzier-language-online-appendix.pdf>

Pamela Jakiela and Owen Ozier, 2019. “Gendered Language.” CGD Working Paper 500. Washington, DC: Center for Global Development. <https://www.cgdev.org/publication/gendered-language>

Center for Global Development
2055 L Street NW
Washington, DC 20036

202.416.4000
(f) 202.416.4050

www.cgdev.org

The Center for Global Development works to reduce global poverty and improve lives through innovative economic research that drives better policy and practice by the world’s top decision makers. Use and dissemination of this Working Paper is encouraged; however, reproduced copies may not be used for commercial purposes. Further usage is permitted under the terms of the Creative Commons License.

The views expressed in CGD Working Papers are those of the authors and should not be attributed to the board of directors, funders of the Center for Global Development, or the authors’ respective organizations.

Contents

1 Introduction	2
2 Grammatical Gender	7
3 Conceptual Framework	13
3.1 Education	14
3.2 Labor Force Participation and the Division of Household Tasks	15
3.3 Implications of the Model	19
4 Data	20
4.1 Building a Grammatical Gender Data Set	20
4.2 Other Sources of Data	22
5 Cross-Country Analysis	23
5.1 Empirical Strategy	23
5.2 Labor Force Participation	24
5.3 Educational Attainment	26
5.4 Gender Attitudes	27
5.5 Robust Inference	28
6 Within-Country Analysis	32
6.1 Empirical Strategy	32
6.2 Results	34
7 Causality	35
8 Conclusion	37
References	39

1 Introduction

Language structures thought. All human beings use language to articulate their ideas and communicate them to others. Yet, the world’s languages show tremendous diversity in terms of their structure and vocabulary. Different languages obviously use different words to describe the same concept, but they also organize the relationships between concepts in remarkably different ways. Because languages are so diverse and language is so fundamental to thought, some scholars have argued that the language we speak may limit the scope of our thinking. Benjamin Lee Whorf, one of the original proponents of this theory of *linguistic determinism*, famously argued that it was difficult for humans to think about ideas or concepts for which there was no word in their language (Whorf 2011[1956]a).

Though specious anecdotes about obscure languages abound, cognitive scientists have largely refuted the strongest forms of Whorf’s hypothesis (Boroditsky, Schmidt, and Phillips 2003). Nonetheless, there is mounting evidence for weaker forms of linguistic determinism: the languages we speak shape our thoughts in subtle, subconscious ways. For example, implicit association tests show that bilinguals display different subconscious attitudes when tested in their different languages (Ogunnaike, Dunham, and Banaji 2010, Danziger and Ward 2010). Differences in language structure also influence our behavior in the economic realm. Chen (2013), for instance, demonstrates that speakers of languages that demarcate the future as separate from the present (e.g. English) save less than those whose languages make no such distinction (e.g. German).

Several recent papers explore the link between language and gender roles. As Alesina, Giuliano, and Nunn (2013) note, views of the appropriate role for women in society differ markedly across cultures. Languages also vary in their treatment of gender. At one extreme, languages such as Finnish and Swahili do not mark gender distinctions in any systematic way: nouns are not categorized as either masculine or feminine; and the same first, second, and third person pronouns are used for males and females. Many languages distinguish between human males and females by using different pronouns: for example, “he” and

“she” in English. Some languages go even further, extending the gender distinction to inanimate nouns through a system of *grammatical gender*. For example, languages such as Spanish and Italian partition all nouns — even inanimate objects — into distinct gender categories. This feature of language forces gender into every aspect of life. For a speaker of a *gender language*, gender distinctions are salient in every thought and utterance: the space of words is divided into distinct masculine and feminine spheres, and one must constantly reference this mental partition to produce grammatically correct speech.

Does grammatical gender shape (non-grammatical) gender norms? Does it impact women’s participation in economic life? Writing nearly 100 years ago, Benjamin Lee Whorf argued that the existence of linguistic gender categories likely made other gender divisions appear more natural (Whorf 2011[1956]b), though he did not provide any empirical evidence that this was the case. However, recent work by social scientists supports his claim. For example, seemingly arbitrary grammatical gender distinctions do influence our subconscious thoughts, imbuing inanimate nouns with masculine or feminine attributes (e.g. strength or beauty) in line with their assigned grammatical gender category (Boroditsky, Schmidt, and Phillips 2003). Pérez and Tavits (forthcoming) show that Estonian/Russian bilinguals are more supportive of gender equality when interviewed in (non-gender) Estonian than in (gender) Russian.

Whether this pattern extends beyond specific cases has been difficult to assess empirically. In the economic realm, one recent study of immigrants to the United States shows that those who grew up speaking a gender language are more likely to divide household tasks along gender lines (Hicks, Santacreu-Vasut, and Shoham 2015), while another demonstrates that female labor supply is lower among immigrants who speak a gender language at home (Gay, Hicks, Santacreu-Vasut, and Shoham 2017). These analyses make use of the most comprehensive existing data source on languages, the World Atlas of Language Structures (WALS). The WALS documents whether a language employs grammatical gender, but only for a fraction of the world’s languages. Using it alone, analysis within Africa or Asia — where widely-spoken indigenous languages differ in their grammatical gender structure —

is nearly impossible. Cross-country analysis using the WALS relies on the assumption that missing data on the native languages of half the world’s population is ignorable, yielding a set of bounding and clustering problems that severely hamper inference.¹ Progress on this research topic demands a new source of data.

We provide new evidence to support the hypothesis that grammatical gender shapes views of women’s role in society and directly impacts women’s labor force participation. To do this, we construct a data set characterizing the grammatical gender structure of 4,346 living languages, expanding the number of languages for which systematic data on grammatical gender is available by almost a factor of ten. We draw on a range of data sources including language textbooks, historical records, academic work by linguists, and — in a small number of cases — firsthand accounts from native speakers and translators; using these data sources, we generate a measure of the grammatical gender structure of each of the languages in our data set. Taken together, these languages account for 6.44 billion people, or over 99 percent of the world population.²

We use these data in two ways. First, we calculate — for every country in the world — an estimate of the proportion of the population whose native language is a gender language. We are able to account for more than 90 percent of the estimated population in all but three countries. In our first piece of analysis, we explore the cross-country relationship between grammatical gender and women’s labor force participation, women’s educational attainment, and gender attitudes among both men and women. We then complement our cross-country analysis by estimating the individual-level association between grammatical gender and women’s participation in economic life in countries where both gender and non-gender languages are indigenous and widely spoken. We do this within-country analysis separately in two contexts: using Afrobarometer data from four African countries (Kenya, Niger, Nigeria, and Uganda) and, separately, using the India Human Development Survey, which covers 33 Indian states.

Our cross-country analysis suggests a robust negative relationship between grammati-

¹WALS has also been used to study origins of language structures, as in Galor, Ozak, and Sarid (2018).

²This calculation is based on *Ethnologue* estimates of the total number of native speakers in the world.

cal gender and female labor force participation. Our preferred specification suggests that grammatical gender is associated with a 12 percentage point reduction in women’s labor force participation and an almost 15 percentage point increase in the gender gap in labor force participation. These associations are robust to the inclusion of a wide range of controls including suitability for the plough. Taken at face value, our coefficient estimates suggest that gender languages keep approximately 125 million women around the world out of the labor force. Following the approach suggested by Altonji, Elder, and Taber (2005) and Oster (2017), we estimate that unobservable country-level characteristics would need to be 1.44 times more correlated with treatment than observed covariates to fully explain the apparent impact of grammatical gender on the level of female labor force participation; unobserved factors would need to be 3.24 times more closely linked to treatment to explain the impact of grammatical gender on the gender gap in labor force participation.

We find a far more muted cross-country relationship between grammatical gender and women’s educational attainment. This may be due to the fact that the average within-country gender gap in educational attainment is much smaller than the gender gap in labor force participation — since many wealthy countries have no gender gap in educational attainment, particularly at the primary school level. The prevalence of gender languages is negatively associated with the gender gap in primary school completion after controlling for continent fixed effects, but the estimated relationship is only marginally statistically significant.

Using data from the World Values Survey (WVS), we show that grammatical gender predicts support for traditional gender roles. The coefficient estimate is large in magnitude, suggesting that differences in language could explain the entire gap in gender attitudes between Ukraine (at the 55th percentile of WVS countries in terms of support for gender equality) and Trinidad and Tobago (at the 80th percentile). As Whorf might have hypothesized, gender languages are associated with greater support for traditional gender roles among both men and women.

Though our analysis uses much richer country-level data on grammatical gender than

has previously been available, mis-measurement of our independent variable of interest is still a potential concern. We would typically expect measurement error in the independent variable to bias the estimated association toward zero, but the interval nature of our measure of the country-level prevalence of grammatical gender (when the gender structure of the native language is not known for the entire population) can also lead to invalid inference. Using a bounding technique proposed by Imbens and Manski (2004), we show that our results are robust to correcting for the censored nature of our independent variable of interest.

A more serious inference concern arises from the fact that languages are not independent. Within a language family, individual tongues have evolved in parallel over many centuries. While this slow process of language development may help to address potential concerns about reverse causality, it complicates statistical inference. Intuitively, language characteristics are assigned at the level of a “cluster,” but countries draw from many different “clusters.” In relation to grammatical gender, these “clusters” do not exactly coincide with any existing categorical tier of language families. We address this issue by implementing a permutation test that respects both the distribution of languages across countries and the observed pattern of variation in treatment (i.e. grammatical gender) across and within language families. We cluster languages at the highest level of the language tree where we do not observe variation in grammatical gender. Generating 10,000 hypothetical assignments of grammatical gender across the 203 clusters so generated allows us to calculate permutation-test p-values indicating the likelihood that the association between grammatical gender and our outcomes of interest would be as strong as the observed relationship under the null hypothesis — given the structure of the language tree, the observed variation in grammatical gender across languages, and the distribution of languages across countries. Results suggest that the strong association between grammatical gender and women’s labor force participation is not spurious.

To further assess the likelihood of a causal link between gender languages and women’s involvement in economic life, we examine the individual-level association between gram-

grammatical gender and women’s labor force participation and educational attainment in two parts of the world where both gender and non-gender languages are indigenous and widely spoken: Sub-Saharan Africa and India. Combining our language data with (i) Afrobarometer surveys from Kenya, Nigeria, Niger, and Uganda and (ii) the India Human Development Survey, we show that — within countries — grammatical gender is associated with larger gender gaps in educational attainment and labor force participation in two distinct contexts. Women whose native language is a gender language obtain less education and are less likely to be in the labor force than women whose native language is not a gender language, even after controlling for interactions between gender (i.e. the indicator for being female) and religious affiliation. The approach suggested by Altonji, Elder, and Taber (2005) and Oster (2017) suggests that unobservable characteristics are unlikely to explain the relationship. Thus, gender languages appear to reduce women’s labor force participation and lower their educational attainment in both Sub-Saharan Africa and India.

The rest of this paper is organized as follows. Section 2 introduces the concept of grammatical gender and surveys recent research on its impacts. Section 3 presents a theoretical framework illustrating the channels through which grammatical gender might lead to larger gender gaps in educational attainment and labor force participation. Section 4 provides an overview of our data sources, including the data we have compiled on the grammatical structure of more than 4,000 languages. Section 5 presents our cross-country analysis, and Section 6 presents individual-level, within-country analysis. Section 7 discusses causality. Section 8 concludes.

2 Grammatical Gender

Many languages partition the set of all nouns into mutually exclusive categories. Membership in these categories, which are typically referred to as either *genders* or *noun classes* (Corbett 1991, Aikhenvald 2003), can be manifest in several ways. Members of a noun class may be semantically related, or they may be linked by morphology. For example, members

of the KI-/VI- class in Swahili often begin with *ki-* in the singular and *vi-* in the plural — e.g. “chair” is *kiti* and “chairs” is *viti*. However, though semantic and morphological regularities are a common characteristic of noun classes, they are not required. Instead, membership in a specific noun class is defined based on agreement: class must be reflected in the conjugation of associated words within the noun phrase or predicate in grammatically correct speech (Aikhenvald 2003).³ In Swahili, for example, the noun class determines the prefixes used to modify adjectives, verbs, demonstratives, and other parts of speech. So, “these new chairs” is *viti vipya hivi*, while “these new teachers” is *walimu wapya hawa* because the word “teacher” is part of the M-/WA- class rather than the KI-/VI- noun class.⁴ Nouns are said to belong to the same agreement class if, “given the same conditions, they will take the same agreement form” (Corbett 1991, p. 148), where the relevant “conditions” are linguistic and typically relate to number and case.

Systems of noun classification differ widely across languages, and not all languages have such a system. One of the most common bases for a system of noun classification is biological sex: (some) female humans and some other nouns are assigned to one category, while (some) male humans and some other nouns are assigned to a different category (Corbett

³There is some debate among linguists as to whether agreement rules that do not involve elements of the noun phrase or the predicate can form the basis of a noun class system — specifically, linguists disagree as to whether requiring “anaphoric agreement” between nouns and associated pronouns constitutes a system of grammatical gender (Corbett 1991, Aikhenvald 2003). Corbett (1991) argues that there is no fundamental distinction between pronominal agreement and other forms of grammatical agreement; he consequently classifies languages that (only) require pronominal agreement (e.g. English) as gender languages in his work (Corbett 2013a, Corbett 2013b, Corbett 2013c). Aikhenvald (2003) agrees that there is no fundamental distinction between pronominal agreement and other forms of grammatical concordance, but advocates the use of the traditional definition of grammatical gender to avoid confusion. She also suggests restricting the use of the term “grammatical gender” to systems of noun classification involving a relatively small number of categories that include masculine and feminine. Since our focus is on the links between grammatical gender and non-grammatical gender norms, we adopt her terminology to avoid confusion. Employing the traditional definition of grammatical gender also facilitates the use of data from a wide range of linguistic and anthropological sources, since many historical sources distinguish between grammatical gender (which involves the assignment of nouns to gender categories) and systems that mark natural/human gender morphologically.

⁴Corbett (1991) states: “The existence of gender can be demonstrated only by agreement evidence...Evidence taken only from the nouns themselves, such as the presence of markers on the nouns, does not of itself indicate that a language has genders (or noun classes); if we accepted this type of evidence, then we could equally claim that English had a gender comprising all nouns ending in *-ion*.” Thus, though many nouns within a class may share particular prefixes or suffixes, it is the requirement that other parts of speech (particularly elements of the noun phrase or the predicate) conjugate or inflect appropriately that distinguishes noun classes from other phonological or orthographic partitions of the set of all nouns.

1991, Aikhenvald 2003, Hellinger 2003).⁵ Following Aikhenvald (2003) and Hellinger and Bußman (2003), we refer to systems which assign nouns, including some inanimate nouns, to agreement classes that are based on biological sex as *grammatical gender*; we refer to languages characterized by such systems of grammatical gender as *gender languages*.⁶ Spanish is a prominent example of a gender language: all Spanish nouns are either masculine or feminine, and both definite articles and adjectives must be consistent with a noun’s gender. So, for example, “the white house” is

<i>la</i>	<i>casa</i>	<i>blanc-a</i>
the.FEM	house	white-FEM,

because “house” is feminine, but “the white horse” is

<i>el</i>	<i>caballo</i>	<i>blanc-o</i>
the.MASC	horse	white-MASC

because “horse” is masculine. A Spanish speaker must therefore maintain a mental map that assigns each noun to one of these two distinct gender categories.

Systems of grammatical gender differ along several dimensions.⁷ Gender languages differ in the extent of agreement across parts of speech, and the extent to which the gender distinction represents a complete partition of the set of all nouns. Languages such as Spanish — with only two sex-based noun classes — are at one end of this spectrum. In such languages,

⁵Almost all languages also distinguish between singular and plural, but this is not typically treated as a system of noun classification because the singular and plural forms are treated as two variants of the same noun.

⁶Swahili, for example, has noun classes which determine agreement, but it is not a gender language because none of the Swahili noun classes relates to biological sex in any way.

⁷Moreover, grammatical gender is only one of several ways that grammatical rules can make human gender distinctions salient. For instance, though typically not classified as a gender language, English employs a system of pronominal agreement — different third-person singular pronouns are used for male and female humans and, in some cases, male and female animals (Aikhenvald 2003, Boroditsky, Schmidt, and Phillips 2003, Hellinger and Bußman 2003, Kilarski 2013). Female pronouns have also traditionally been used to refer to ships and other large transportation vessels. Because pronouns agree with the natural gender of animate nouns, Corbett (1991) classifies English as a gender language with a strictly semantic system of noun classification (i.e. a system of grammatical gender based only on biological gender). Such systems of pronominal agreement based on the biological gender of animate referents (rather than the grammatical gender of the nouns themselves) are present in many languages that show no other form of gender inflection (Aikhenvald 2003, Creissels 2000). Other languages — e.g. Finnish, Hungarian, and Swahili — make no grammatical distinction between males and females. Givati and Troiano (2012) show that countries where the dominant language makes pronominal gender distinctions have shorter government-mandated maternity leaves.

every inanimate noun *must* be classified as either feminine or masculine. Languages such as German display a weaker form of grammatical gender because some objects are classified as neither feminine nor masculine. Intuitively, one might think that the partition of nouns into two dichotomous genders suggests that other aspects of the universe should also be so organized (for example, into male and female household tasks). In systems that assign objects (i.e. nouns) without natural gender to gender categories, there is also the question of what the observed grouping signals about the relative status of women and men. Though the rules used to assign nouns to different classes are often phonological (e.g. Spanish nouns that end in “o” are typically masculine), many languages assign some nouns to the feminine gender using semantic guidelines that have a certain cultural intelligibility. For example, dangerous objects are feminine in the Australian language Dyirbal (Lakoff 1987), while one linguist studying the Siberian language Ket suggested that certain small animals were feminine “because they are of no importance to the Kets” (Corbett 1991, p. 19).^{8,9}

⁸In many languages, the grammatical gender of inanimate objects reflects stereotypes about the physical distinctions between males and females. For example, in his discussion of the major Indo-Aryan languages (Bengali, Gujarati, Hindi, Marathi, Oriya, Panjabi, and Sindhi), John Beames (1875) notes: “In all the five languages which have gender expressed, the masculine is used to denote large, strong, heavy, and coarse objects; the feminine weak, small, and fine ones” (p. 148). In the Papuan language Manangu, inanimate objects that are long or thin are masculine, while those that are short or round are feminine (Aikhenvald 2003).

⁹No one knows exactly why grammatical gender systems arose in some language families and not in others. Janhunen (1999) hypothesizes that a single innovation in an ancient West Asian language brought grammatical gender into the Indo-European language family, but grammatical gender arose in indigenous language families on every continent. It is, of course, impossible to fully rule out the possibility that some aspect of culture contributed to the emergence of grammatical gender in certain ancestral languages. That said, since language structures evolve over centuries, even millennia, present-day gender attitudes cannot have had a causal impact on modern grammatical structures. Moreover, we have a relatively good understanding of the process through which grammatical gender was lost from certain widely spoken Indo-European languages; this evidence does not suggest a causal relationship between gender norms and the loss of grammatical gender. For example, McWhorter (2005) argues that the influx of Scandinavian adults into the community of English speakers contributed to the loss of grammatical gender, as an imperfect grasp of inflectional agreement paradigms is common among non-native speakers. This “contact hypothesis” may also explain why grammatical gender is typically absent from Creole languages (McWhorter 2005, Muhleisen and Walicek 2010). However, the reduction and simplification of languages resulting from an influx of non-native speakers is not restricted to the loss of grammatical gender (and has no inherent relationship to societal gender norms): McWhorter (2005) argues that the contact hypothesis also explains why Swahili is one of the few Bantu languages that is not tonal. Kastovsky (1999) proposes a complementary explanation, arguing that the English case-number-gender agreement system was, in essence, made precarious by its own complexity and the absence of reliable morphological rules that could be used to predict agreement classes; in this context, small changes in pronunciation could lead to the conflation of declensional paradigms and their subsequent loss. Aikhenvald (2003) points to a similar process of declensional conflation and subsequent gender loss in Bengali and Persian, and to a parallel loss of the neuter gender in French. Thus, the existing

Whether grammatical gender distinctions influence (non-grammatical) gender attitudes is an empirical question, but the idea that they might is not new. Whorf, for example, argued that gender distinctions in language might make a gendered division of labor seem more natural, suggesting that viewing the world through the lens of a gender language would create “a sort of habitual consciousness of two sex classes as a standing classificatory fact in our thought-world” (Whorf 2011[1956]b, p. 69).¹⁰ This argument — which Whorf advanced without offering any empirical evidence to support it — has been controversial, to say the least. However, recent work in psychology and political science shows that grammatical gender shapes our subconscious attitudes in subtle and surprising ways. For example, Boroditsky, Schmidt, and Phillips (2003) conducted a study — in English — of native speakers of Spanish and German (all of whom were fluent in English); participants in the study were asked to provide (English) adjectives to describe pictures of objects that had been chosen because they had opposite grammatical genders in Spanish and German. Subjects tended to choose adjectives that aligned with the grammatical gender of the noun in their native language. For example, native German-speakers described a picture of a bridge (which is feminine in German) as “beautiful” and “elegant” while native Spanish-speakers described the same (masculine in Spanish) bridge as “big” and “dangerous” (Boroditsky, Schmidt, and Phillips 2003). Thus, the results suggest that grammatical gender shapes the way we think about inanimate objects without inherent biological gender. Grammatical gender also appears to shape gender attitudes — even within individuals. Pérez and Tavits (forthcoming) conduct a survey experiment with Estonian/Russian bilinguals, randomizing the language in which they are interviewed. They show that bilinguals who are interviewed in Russian (a gender language) are less supportive of gender equality than those who are interviewed in (non-gender) Estonian, even though interview languages

evidence tends to suggest that grammatical gender is most often lost through an interplay between linguistic factors (e.g. sound change, similarity between agreement paradigms) and the arrival of large numbers of non-native speakers within a linguistic community.

¹⁰His argument echoes earlier work by Durkheim and Mauss (1963), who highlighted the parallels between culture-specific systems for classifying humans and those used for classifying other aspects of reality. Describing the extension of the clan system of one group of native Australians to the universe of animals and inanimate objects, they wrote: “The reasons which led to the establishment of the categories have been forgotten, but the category persists and is applied, well or ill, to new ideas” (p. 21).

were randomly assigned.¹¹

Recent work also suggests that the influence of grammatical gender extends into the economic realm. Using the World Atlas of Language Structures (WALS), a comprehensive data set on the grammatical structure of more than 500 languages, a number of authors have examined the links between grammatical gender and economic and political outcomes. For example, Mavisakalyan (2015) and Shoham and Lee (2017) use the WALS to examine the cross-country association between grammatical gender and gender inequality in the labor force. Santacreu-Vasut, Shoham, and Gay (2013) show that countries where the national language uses a sex-based system of grammatical gender are less likely to implement gender quotas for political office, while Santacreu-Vasut, Shenkar, and Shoham (2014) find that those countries also have relatively fewer women in corporate leadership positions. Hicks, Santacreu-Vasut, and Shoham (2015) show that immigrants to the United States assign tasks within the household along gendered lines if they grew up speaking a gender language; no such difference is found among immigrants who came to the U.S. before the age of language acquisition, or among the children of immigrants.¹² Importantly, these findings suggest that one’s native language plays a particularly crucial role in shaping one’s views on the appropriate role for women in society.

These analyses suffer from the incompleteness of the WALS. Using it alone, within-country analysis of data from Africa or Asia is not feasible. Cross-country regressions require researchers to calculate country-level averages of a variable (the grammatical gender structure of one’s native language) that is missing for half the world’s population. One of the most cautious approaches to this missing data problem is to use Manski-style bounds,

¹¹There is also evidence that pronominal gender impacts the salience of gender distinctions. Guiora (1983) finds that children who grow up speaking Hebrew, English, or Finnish come to understand their own biological genders at different ages; those who grow up using different pronouns for males and females become aware of their own natural gender earlier. As discussed above, English has a system of pronominal gender while Finnish does not. Hebrew also uses a dichotomous system of grammatical gender (all nouns are either masculine or feminine), and male and female Hebrew-speakers must use grammatically correct verb forms, for example, that reflect their natural gender. Hebrew also uses different second-person pronouns for males and females.

¹²In related work, Gay, Hicks, Santacreu-Vasut, and Shoham (2017) find that female immigrants to the United States exhibit lower labor market participation (working fewer hours, fewer weeks, etc.) if they speak a gender language at home.

but doing so yields upper and lower bounds which contain almost the entire support of conceivable values. Moreover, the absence of data also limits the extent to which one can correct for the non-independence of languages within families while maintaining adequate statistical power.¹³ Robust inference requires an expanded data set on linguistic structures.

3 Conceptual Framework

Existing work examining the empirical relationship between grammatical gender and women’s involvement in economic life has not formally specified the potential causal pathway. In this section, we outline a stylized model that illustrates how grammatical gender — which may predispose us to think of things as either masculine or feminine — could induce gender disparities in education and labor force participation. The model is inspired by Whorf’s suggestion that a grammatical gender system makes the partition of the non-linguistic world into masculine and feminine domains appear more natural. We formalize this intuition by introducing a psychic cost $\phi > 0$ that a person who has grown up speaking a gender language experiences when she (resp. he) enters a domain dominated by the opposite sex.

We endogenize the definition of masculine and feminine domains by assuming that a domain (e.g. a school, the workforce, etc.) is masculine (resp. feminine) whenever the proportion of women (resp. men) in that domain falls below some threshold $\lambda \in [0, 1]$. Thus, when the proportion of women in, say, the workforce is below λ , the work world is perceived as a masculine domain — so, women face a psychic cost when they choose to work outside the home. Symmetrically, if the proportion of women in the workforce exceeds $1 - \lambda$, the workforce would be perceived as a feminine domain, and men would face a psychic cost when they chose to work. Equilibrium requires that each individual make a rational choice about whether or not to enter a domain conditional on the cost structure that results from the realized distribution of genders across each domain.¹⁴

¹³We discuss these clusters further in Section 5.5.2.

¹⁴To focus on the key implications of the model, we assume that those who did not grow up speaking gender languages do not experience such psychic costs — though, of course, they may experience other social or other emotional costs when entering environments where they do not fit in. One could easily extend our

3.1 Education

We consider a simple model of educational attainment where students attend school whenever the expected benefits exceed the immediate costs. The net return to education depends on ability and may also differ across genders. We formalize the set-up as follows, first without grammatical gender and then introducing it. Girl i 's ability is given by $\gamma_i > 0$, where $\gamma \sim F_\gamma$ (for some smooth, etc. function F_γ). Let $R_g(\gamma_i)$ denote the net return to schooling for a girl with ability level γ_i . Without loss of generality, we assume that $R_g(\cdot)$ is net of any monetary costs of attending school. The return to education is increasing in ability: $R'_g(\gamma_i) > 0$. In the absence of grammatical gender, a girl will attend school whenever $R_g(\gamma_i) > 0$. As a result, there exists γ^* such that $R_g(\gamma^*) = 0$, and a proportion $1 - F_\gamma(\gamma^*)$ of girls (all those with $\gamma_i \geq \gamma^*$) attend school.

The setup is symmetric for boys. Boy i 's ability is given by $\beta_i > 0$, where $\beta \sim F_\beta$. In the absence of grammatical gender, a boy with ability level β_i will attend school whenever $R_b(\beta_i) > 0$. There exists β^* such that $R_b(\beta^*) = 0$, and all boys with $\beta_i \geq \beta^*$ attend school. With equal numbers of girls and boys in the population, girls represent proportion

$$P_{girls}^* = \frac{1 - F_\gamma(\gamma^*)}{2 - F_\beta(\beta^*) - F_\gamma(\gamma^*)} \quad (1)$$

of students enrolled in school. The model is symmetric: if $F_\gamma = F_\beta$ and $R_g(\cdot) = R_b(\cdot)$, then $\lambda^* = \frac{1}{2}$ and $\gamma^* = \beta^*$.

When grammatical gender predisposes us to view domains as either masculine or feminine, there are three possible equilibria: school can be either masculine, neutral (non-gendered), or feminine. In the masculine equilibrium (if it exists), boys attend school whenever $R_b(\beta_i) \geq 0$, but girls only attend if $R_g(\gamma_i) \geq \phi$ — for girls, going to school entails a psychic cost because they perceive school as a masculine domain. An equilibrium exists if the set of children who would attend school conditional on the distribution of psychic costs associated with that equilibrium yields a gender composition (of students)

model to consider the possibility that these costs exist for everyone but might be larger when partitioning the world in masculine, feminine, and potentially neutral spheres appears more natural.

consistent with that equilibrium. So, for example, it is possible for school to be a masculine domain in equilibrium if the set of students who would attend school when girls face a psychic cost but boys do not skews sufficiently male to keep the proportion of girls in the student body below λ .

As we show in the Online Appendix, at least one of the three possible equilibria always exists. More interestingly, multiple equilibria are often possible, but both welfare and human capital attainment are highest in the gender-neutral equilibrium. In this context, policies such as single-sex schools can improve welfare and increase human capital by allowing girls (or boys) to attend school without the psychic costs associated with entering an environment that is perceived as the domain of the opposite sex.¹⁵ Other policies that increase the net return to education — for example, eliminating school fees or making education compulsory (which introduces costs for non-attendance) — can have indirect effects on female enrollment by changing the expected proportion of girls who attend school. If these policies bring the expected ratio of girls to boys closer to parity, the gendered equilibrium may cease to exist. Moreover, when multiple equilibria are possible, such policies have the potential to nudge a society from one feasible equilibrium to another.

3.2 Labor Force Participation and the Division of Household Tasks

Next, we consider the decision problem facing two parents who maximize their consumption while caring for their children. Again, we assume that the ability of girl/female/woman/mother i is characterized by $\gamma_i \sim F_\gamma$ and the ability of boy/male/man/father is characterized by $\beta_i \sim F_\beta$.¹⁶ γ and β both have continuous support between 0 and some finite maxima, β^{max} and γ^{max} .

A household maximizes consumption:

$$C = w_{mom}L_{mom} + w_{dad}L_{dad} - w_{nanny}H_{nanny} \quad (2)$$

¹⁵Fryer and Levitt (2010) discuss the prevalence of same-sex schools in middle Eastern countries where gender gaps in educational attainment are small but gender gaps in labor force participation persist.

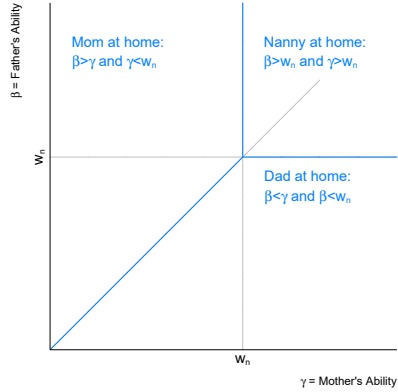
¹⁶For obvious reasons, using the subscripts m and f to distinguish between male and female adults who are also mothers and fathers might be confusing.

where $w_{mom} = \gamma_i$ indicates the wage that a mom of ability γ_i earns if she works outside the home, $w_{dad} = \beta_i$ the wage that a dad of ability β_i earns if he works outside the home, and w_{nanny} represents the market wage paid to nannies. We assume that nannies are female, and that they are young women who would not be included in the adult labor force if they were not employed as nannies (for example, au pairs, older sisters).¹⁷ Both mom and dad have one unit of time which they allocate to either work outside the home or childcare: $H_{mom} + L_{mom} = 1$ and $H_{dad} + L_{dad} = 1$. One unit of adult time must be spent caring for the child: $H_{mom} + H_{dad} + H_{nanny} = 1$.

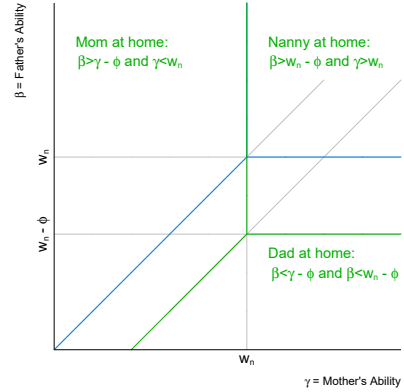
First, consider the case where there are no gendered domains. A household will hire a nanny to take care of the children whenever both the mother and the father are both able to earn more than the nanny's wage — i.e. when $\beta_i \geq w_n$ and $\gamma_i \geq w_n$. When $\gamma_i < w_n$ and $\gamma_i \leq \beta_i$, the mother stays home while the father works. When $\beta_i < w_n$ and $\gamma_i > \beta_i$, the father stays home while the mother works. Panel A of Figure 1 illustrates this partition of the space of possible parental ability levels.

Figure 1: Labor Force Participation in Two Equilibria

Panel A: Domains Not Gendered



Panel B: Home Is Feminine



¹⁷While this assumption is realistic in a range of contemporary and historical settings, it also serves a purpose by increasing the likelihood that the home environment is a predominantly a feminine domain. Other ways of achieving the same goal (for example, endogenizing fertility and making it costly for women to enter the work force when children are very young) make the model too realistic to be useful.

Let $f_{\beta,\gamma}(\beta, \gamma)$ denote the joint distribution of β and γ . In the absence of gendered domains, define P_{mom}^* as the proportion of households where the mother stays at home:

$$P_{mom}^* = \int_{\beta=0}^{\beta=w_n} \int_{\gamma=0}^{\gamma=\beta} f_{\beta,\gamma}(\beta, \gamma) + \int_{\beta=w_n}^{\beta=\beta^{max}} \int_{\gamma=0}^{\gamma=w_n} f_{\beta,\gamma}(\beta, \gamma). \quad (3)$$

In other words, P_{mom}^* is the integral of $f_{\beta,\gamma}(\beta, \gamma)$ over the “mom at home” region in Figure ???. P_{dad}^* and P_{nanny}^* can be defined analogously:

$$P_{dad}^* = \int_{\beta=0}^{\beta=\gamma} \int_{\gamma=0}^{\gamma=w_n} f_{\beta,\gamma}(\beta, \gamma) + \int_{\beta=w_n}^{\beta=\beta^{max}} \int_{\gamma=w_n}^{\gamma=\gamma^{max}} f_{\beta,\gamma}(\beta, \gamma) \quad (4)$$

and

$$P_{nanny}^* = \int_{\beta=w_n}^{\beta=\beta^{max}} \int_{\gamma=w_n}^{\gamma=\gamma^{max}} f_{\beta,\gamma}(\beta, \gamma). \quad (5)$$

For any $f_{\beta,\gamma}(\beta, \gamma)$, $P_{mom}^* + P_{dad}^* + P_{nanny}^* = 1$ since households must either have mom, dad, or a nanny at home with the children. Since all households have exactly one person at home, the proportion of homes where a woman takes care of the children is $P_{mom}^* + P_{nanny}^*$. The proportion of women in the (out-of-the-home) workforce is:

$$\frac{P_{dad}^* + P_{nanny}^*}{1 + P_{nanny}^*} \quad (6)$$

since households with a nanny at home send both a man and a women into the workforce. $(P_{mom}^*, P_{dad}^*, P_{nanny}^*)$ is an equilibrium in a trivial sense, since every household optimizes and individual (household) optima are not strategically interdependent.

When individuals are predisposed to view domains as gendered (so λ and ϕ play a role in decision-making), the equilibrium described above is one of nine that might exist. Home and work can both be either masculine, , or neutral (non-gendered), or feminine. Each of the nine candidate equilibria is a pair HW where $H \in \{M, N, F\}$ characterizes the ‘home’ domain and $W \in \{M, N, F\}$ characterizes the “work” domain. So, the NN equilibrium would be one in which neither home nor work is a gendered domain, whereas the FM equilibrium would be one in which home is a feminine domain and work is a masculine

domain.

The NN equilibrium, if it exists, is characterized by the same pattern of observed in the absence of grammatical gender (as shown in Panel A of Figure 1): both parents work whenever $\gamma_i > w_n$ and $\beta_i > w_n$, and the parent who would earn the lower wage stays home with the child otherwise. Hence, $P_{dad}^{NN} = P_{dad}^*$, $P_{mom}^{NN} = P_{mom}^*$, and $P_{nanny}^{NN} = P_{nanny}^*$. However, when domains can be gendered, this is only an equilibrium when

$$\lambda < P_{mom}^{NN} + P_{nanny}^{NN} < 1 - \lambda \quad (7)$$

and

$$\lambda < \frac{P_{dad}^{NN} + P_{nanny}^{NN}}{1 + P_{nanny}^{NN}} < 1 - \lambda. \quad (8)$$

In other words, the equilibrium proportion of women taking care of children (i.e. households where a female takes care of the child) and the proportion of women in the (out-of-the-home) workforce must both fall between λ and $1 - \lambda$ for a neutral equilibrium — in which neither home nor work is a gendered domain — to exist. It is apparent that this becomes less likely when λ is close to one half and the scope for non-gendered domains is limited.

Next, consider the FN equilibrium, where home is a feminine domain but work is neither masculine nor feminine. If such an equilibrium exists, a man who stays home with his children will experience a psychic cost of $\phi > 0$. Total household utility if the father provides childcare is therefore given by

$$C = \gamma_i - \phi. \quad (9)$$

In this equilibrium, a household will hire a nanny whenever $\beta_i > w_n - \phi$; the father will stay home whenever $\beta_i < \gamma_i - \phi$ and $\beta < w_n - \phi$; and the mother will stay home whenever $\beta_i > \gamma_i - \phi$ and $\gamma_i < w_n$. As Panel B of Figure 1 illustrates, two types of men who do not work in the NN equilibrium will work in the FN equilibrium. Men whose wives work (because $\gamma_i > w_n$) will now enter the workforce whenever their ability (β_i) falls between

$w_n - \phi$ and w_n . Men will also work whenever $\gamma_i - \phi < \beta_i < \gamma_i < w_n$; their higher-ability wives will stay home because childcare (“women’s work”) entails a psychic cost for men when relatively few men stay home. Both of these changes lower the average ability level among those in the labor force.

In the Online Appendix, we characterize the feasible equilibria in greater detail and demonstrate that at least one equilibrium always exists. As in the case of educational attainment, multiple equilibria are possible, and the ability level of the labor force is always highest in the NN equilibrium, where neither home nor work is a gendered domain.

3.3 Implications of the Model

The model does not demonstrate that grammatical gender necessarily predicts lower educational attainment and labor force participation among women than men. Instead, we formalize Whorf’s intuition that grammatical gender predisposes us toward the view that men and women should exist in separate domains. We have kept the model as symmetric as possible while still recognizing the empirical fact that women and girls do all of the birthing and most of the childcare work in every human society ever studied (Lancy 2015). Because of its symmetry, our model allows for the possibility that gendered equilibria could exist in which boys attain less education than girls and in which men are less likely to work outside the home than women — though such equilibria are unlikely when the returns to education are higher for males and childcare is costly.

The key prediction of the model is that the minimum equilibrium level of girls’ educational attainment and women’s labor force participation is lower with grammatical gender than without. This motivates the empirical tests presented in the rest of the paper. However, the model also demonstrates that grammatical gender is more a nudge than a constraint. When $\lambda < 0.5$, gender-neutral equilibria become possible.¹⁸ Hence, many soci-

¹⁸There are several ways of making this precise. For example, the closer λ is to zero, the wider the range of returns to education, wages, and joint distributions of ability that support a gender-neutral equilibrium; when ability levels and the returns to education do not differ by gender, any value of $\lambda < 0.5$ permits a gender-neutral equilibrium for education. The situation in the labor market is more complicated, both because of the presence of nannies and because of matching in the marriage market, upon which we have

eties where women’s labor force participation is still very low have the potential to move to a more equitable equilibrium very rapidly. The model also highlights the potential for policy responses that work around the subtle influence of the tendency to partition the world into sex-specific domains by creating female-centric spaces in the modern sector – as many Middle Eastern countries have done to improve girls’ educational outcomes.

4 Data

We compile a new data set characterizing the gender structure of more than 4,000 living languages. Together, the languages that we classify account for over 99 percent of the world’s population. As discussed below, we collate data from a range of academic publications, pedagogical materials (e.g. language textbooks), and historical sources. The downside of this approach is that there may be measurement error at the language level: while many sources explicitly state that a language either does or does not use a system of grammatical gender, we cannot always be certain that the same precise definition of grammatical gender is being used across sources.¹⁹ The strength of our approach is that we are able to characterize the grammatical structure of thousands of languages accounting for almost all of the world’s population.

4.1 Building a Grammatical Gender Data Set

Data on the world’s native languages comes from the *Ethnologue*, a comprehensive database of over 7,000 languages (Lewis, Simons, and Fennig, eds., 2016). Combining the *Ethnologue* data with information on the grammatical gender structure of the world’s languages allows us to construct an estimate of the fraction of each country’s population that speaks a gender language as their native language. Of the 7,457 languages included in the *Ethnologue* database, we drop languages that are extinct or have no native speakers, sign languages,

not imposed any structure.

¹⁹Indeed, even recent work by linguists does not always agree on the definition of grammatical gender — see Corbett (1991) and Aikhenvald (2003) for discussion.

and dying languages that had fewer than 100 native speakers when last assessed by *Ethnologue* researchers. This leaves 6,190 languages. Together, these languages account for an estimated 6.50 billion native speakers. Of these, we successfully identify academic or historical sources characterizing the gender structure of native languages accounting for 6.44 billion native speakers (or more than 99 percent of the total).

Data on the gender structure of languages comes from a range of sources. Three of the best known are: the World Atlas of Language Structures (WALS), which characterizes the noun classification system of 525 languages; George L. Campbell’s *Compendium of the World’s Languages* (Campbell 1991); and George Abraham Grierson’s eleven-volume *Linguistic Survey of India* (Grierson 1903a, 1903b, 1904, 1905, 1907, 1908, 1909, 1916, 1919, 1921), which was compiled between 1891 and 1921 and covers more than 300 South Asian languages and dialects. Additional data on the grammatical gender structures of languages comes from academic articles and teaching materials focused on individual languages. We also collected first-person accounts from native speakers for a small number of relatively undocumented languages (e.g. Fiji Hindi and Rohingya). Detailed information on the range of sources (including quotes used to characterize each language’s grammatical gender) is provided in our (Online) Data Construction Appendix.

For each mother tongue in the *Ethnologue* database, we code two variables characterizing the language’s grammatical gender structure. First, we create an indicator for using any system of grammatical gender. We code a language as a gender language if it meets two criteria: first, the language must use a system of noun classes that includes masculine and feminine as two of the possible categories; second, the masculine and feminine categories must include some inanimate objects — i.e. assignment to the gender noun classes should not be based exclusively on the biological sex (or human gender) of the referents.²⁰ Second, whenever possible, we also code an indicator for dichotomous gender languages

²⁰As discussed above, linguistic sources do not always use the same implicit definition of grammatical gender. For example, the phrase “marks gender” can be used to indicate either grammatical gender or a more limited system of indicating the gender of a human referent. Since many linguistic sources explicitly distinguish between grammatical gender and lexical marking of human/animate gender, we only use sources that indicate whether inanimates are classed in terms of nominal gender.

(e.g. Spanish) that assign all nouns to either the masculine or the feminine noun class.

We successfully classify 4,346 languages which together account for more than 99 percent of the world’s population. We classify all but four of the 383 languages with more than one million native speakers, and we are able to confirm the gender structure using two independent data sources for 324 of these large languages. We are able to account for more than 99 percent of the population in 171 of 193 countries, and we account for less than 95 percent of the population in only eight countries: Eritrea (94.5 percent of native speakers coded), the Iran (93.7 percent), Ethiopia (92.6 percent), the Laos (90.2 percent), Timor-Leste (90.0 percent), Cameroon (89.1 percent), Chad (75.4 percent), and Papua New Guinea (32.0 percent).

Figure 2 characterizes the distribution of gender languages around the world. While many countries are dominated by either gender or non-gender languages, there is considerable within-country variation in Canada and the United States, Sub-Saharan Africa, South Asia, and the Andean region of South America. Across all countries, we estimate that approximately 38.6 percent of the world’s population speaks a gender native language.

4.2 Other Sources of Data

Additional data for our cross-country analysis comes from several sources. Data on labor force participation, income, and population come from the World Bank’s World Development Indicators database. We use data on labor force participation in 2011, which is available for 178 countries. We also use data on primary and secondary school completion from the Barro-Lee Educational Attainment Data Set (Barro and Lee 2013), which is available for 142 countries. Data on gender attitudes comes from the World Values Survey and is available for 56 countries (World Values Survey Association 2015). Finally, we take several country-level geographic controls (average precipitation and rainfall plus suitability for the plough) from Alesina, Giuliano, and Nunn (2013). These data are available for 173 countries.

Data for our individual-level analysis comes from two sources. For African countries,

we use the nationally-representative Afrobarometer Surveys (Afrobarometer Data 2016). Afrobarometer surveys have been conducted in 36 African countries and are representative of the voting age population within each country. We use data from four countries where gender and non-gender languages are indigenous and widely spoken: Kenya, Niger, Nigeria, and Uganda. Data for Niger is only available in Round 5 of the Afrobarometer (2011–2013). For the other three countries, four rounds of data are available: 2002–2003, 2005–2006, 2008–2010, and 2011–2013.²¹ We successfully classify the grammatical gender structure of the native languages of 99.1 percent of respondents, yielding a data set of 26,546 respondents who speak 175 different native languages.

We replicate our within-country analysis for India using the India Human Development Survey (Desai, Dubey, and Vanneman 2015). The IHDS includes data on 76,351 household heads and their spouses living in 33 Indian states. We are able to classify the grammatical gender structure of the native language of 99.5 percent of IHDS respondents, yielding a data set of 75,966 observations.

5 Cross-Country Analysis

5.1 Empirical Strategy

In our cross-country analysis, we examine the association between women’s labor force participation and the proportion of a country’s population whose native language is a gender language, $Gender_c$. Our main empirical specification is an OLS regression of the form:

$$LFP_c = \alpha + \beta Gender_c + \delta_{continent} + \lambda X_c + \varepsilon_c \quad (10)$$

where LFP_c is women’s labor force participation in country c (in 2011), $Gender_c$ is the proportion of the population of country c whose native language is a gender language, $\delta_{continent}$ is a vector of continent fixed effects, X_c is a vector of country-level geography

²¹Kenya, Nigeria, and Uganda were also included in the first round of the Afrobarometer. However, that data set does not contain detailed information on native languages.

controls, and ε_c is a conditionally mean-zero error term.²² Standard errors are clustered at the language level (by the most widely spoken language within each country).

Our main outcome of interest is women’s labor force participation. However, we do not wish to conflate gender differences in labor market participation with structural factors that impact labor force participation among both men and women. To rule out this possibility, we include specifications where the outcome variable is the gender difference in labor supply, i.e. women’s labor force participation minus men’s labor force participation.²³

We also examine two other outcome variables related to gender norms: women’s educational attainment and gender attitudes. Our analysis of educational outcomes parallels our analysis of labor force participation. We examine rates of primary and secondary school completion among women and differences between women’s and men’s completion rates. In our analysis of WVS data on gender attitudes, we construct an index of gender attitudes by taking the first principal component of the eight WVS questions on gender roles. Since we are considering attitudes rather than behaviors, we do not report gender differences; instead we compare attitudes by gender to test whether grammatical gender shapes the views of traditional gender roles among both men and women.

5.2 Labor Force Participation

Figure 3 summarizes female labor force participation in the 178 countries for which data is available. The figure highlights the fact that women’s participation in economic life varies tremendously across countries: the women’s labor force participation rate ranges from 9 percent in the Republic of Yemen to 87 percent in Madagascar. Figure 3 suggests a negative relationship between the prevalence of gender languages and women’s involvement in the labor force. In the figure, darker bars indicate a higher prevalence of grammatical

²²As discussed further below, our results are also robust to the inclusion of additional contemporaneous controls such as log GDP per capita and population. However, such controls might be directly impacted by gender norms and women’s involvement in the labor force, creating a “bad controls” problem and biasing the coefficient of interest (Angrist and Pischke 2008, Acharya, Blackwell, and Sen 2016). We therefore focus on geographic controls — proportion tropical, precipitation, temperature, suitability for the plough, and an indicator for being landlocked — which are plausibly exogenous.

²³As a robustness check, we report specifications that use the ratio of women’s labor force participation to men’s labor force participation as the outcome variable (see Online Appendix Table A1).

gender. It is clear that many of the countries with the lowest levels of women’s labor force participation and the largest gender gaps in labor force participation are those where gender languages are dominant.

We confirm the statistical significance of this relationship in a regression framework in Table 1. In the first three columns, the outcome variable is the average level of female labor force participation in country c . We report a parsimonious specification with no controls in Column 1. Gender languages are negatively and significantly associated with lower levels of female labor force participation. The coefficient estimate suggests that women’s labor force participation is 13.83 percentage points higher in the absence of gender languages (p-value 2.29×10^{-6}). Column 2 of Table 1 reports a specification that includes continent fixed effects; Column 3 also includes geographic controls (percentage tropical, average temperature and precipitation, an indicator for being landlocked, and suitability for plough agriculture). The coefficient of interest is negative and statistically significant in both specifications. Moreover, it remains reasonably similar in magnitude: when all of our geographic controls are included, the coefficient suggests that grammatical gender is associated with an 11.92 percentage point decline in women’s labor force participation (p-value 5.04×10^{-4}).

In Columns 4 through 6 of Table 1, we replicate our analysis using the gender difference in labor force participation as the dependent variable. Gender languages are also associated with robust differences in women’s labor force participation relative to men.²⁴ In a parsimonious specification with no controls (Column 4), we find that grammatical gender is associated with an 11.61 percentage point increase in the gender gap in labor force participation (p-value 6.22×10^{-6}). When we include continent fixed effects and country-level geography controls, the coefficient rises to suggest that grammatical gender is associated with a 14.66 percentage point increase in the gender difference in labor force participation (p-value 1.37×10^{-5}). Thus, the proportion of a country’s population whose native language is a gender language is a robust predictor of gender differences in labor force participation.²⁵

²⁴As shown in Online Appendix Table A1, we obtain similar results when we use the ratio of female labor force participation to male labor force participation as the outcome variable.

²⁵In the Online Appendix, we report a range of robustness checks, all of which suggest that the relationship

Taken at face value, our coefficient estimates suggest that grammatical gender might keep as many as 125 million women around the world out of the labor force.

5.3 Educational Attainment

Next, we examine the association between grammatical gender and women’s educational attainment. Education is a key determinant of wages; in many countries, gender differences in educational attainment translate into gender gaps in wages and economic empowerment (Grant and Behrman 2010). Nonetheless, gender gaps in primary and secondary school completion are not nearly as large as gender gaps in labor force participation. Across the 142 countries in the Barro-Lee data set, the average gender gap in primary school completion is only six percentage points and the average gender gap in secondary school completion is only four percentage points. This reflects the very high rates of primary school completion in many parts of the world: more than two thirds of the countries in the Barro-Lee data set have rates of primary school completion above 90 percent for both men and women. Moreover, many wealthy countries have compulsory schooling laws which reduce gender gaps in educational attainment.

In Table 2, we examine the cross-country relationship between grammatical gender and primary school completion. As expected, the relationship is *positive* and significant when continent controls are not included — reflecting the fact that primary school completion rates are highest in Europe, where gender languages are dominant. Once continent fixed effects are included, the estimated association is negative but not statistically significant. In Columns 4 through 6 of Table 2, we examine the relationship between grammatical gender

between grammatical gender and female labor force participation is not driven by outliers or specification choices. In Online Appendix Table A2, we show that our main result is robust to the inclusion of a range of “bad controls” — intermediate outcomes that could themselves have been impacted by grammatical gender. As is well known, including such controls could bias the coefficient of interest, making it impossible to interpret (Angrist and Pischke 2008, Acharya, Blackwell, and Sen 2016). Nevertheless, we note that our main result is robust to the inclusion of controls for log GDP per capita, population, major world religions, and an indicator for post-Communist regimes. In Online Appendix Table A3, we demonstrate that our results hold when we drop each of the major world languages — Arabic, English, and Spanish. Finally, in Online Appendix Table A4, we include an additional variable for the proportion of a country’s population whose native language is a dichotomous gender language with only two noun classes (masculine and feminine). Results suggest that even weak forms of grammatical gender predict women’s (lack of) involvement in the labor force.

and the gender gap in primary school completion. After including continent fixed effects, we find a negative relationship that is marginally statistically significant. Coefficient estimates suggest that grammatical gender is associated with a 3.72 percentage point increase in the gender gap in primary school completion (Table 2, Column 6, p-value 0.088).

We observe an even more muted cross-country relationship between gender languages and secondary school completion (Table 3). After including continent fixed effects, the association between grammatical gender and female secondary school completion is never statistically significant, nor do we observe a statistically significant association between grammatical gender and the gender gap in secondary school completion. Thus, grammatical gender explains cross-country variation in female labor force participation, but does not explain most of the observed cross-country variation in women’s educational attainment.

5.4 Gender Attitudes

Our main measure of gender attitudes is an index that we construct by taking the first principal component of the eight World Values Survey (WVS) questions related to gender. In Figure 4, we plot the cross-country relationship between each of these questions and the proportion of a country whose native language is a gender language. The prevalence of gender languages predicts responses to seven of the eight WVS questions.

In Table 4, we confirm the association between the prevalence of gender languages and our summary index of gender attitudes in a regression framework. After controlling for continent fixed effects and country-level geography, the coefficient estimate suggests that grammatical gender is associated with greater support for traditional gender roles. To put the coefficient magnitudes in context, the estimates indicate that grammatical gender alone could explain the gap in gender attitudes between Ukraine (at the 55th percentile) and Trinidad and Tobago (at the 80th percentile). Thus, the estimated association between grammatical gender and non-grammatical gender attitudes is both statistically and culturally significant.

If grammatical gender shapes gender attitudes, we would expect it to impact the beliefs

of both men and women. In Table 5, we show that — as expected — we observe a negative association between the country-level prevalence of grammatical gender and gender attitudes among both women (Columns 1 through 3) and men (Columns 4 through 6). The association is always statistically significant after including continent fixed effects. Moreover, though the coefficient is slightly larger for men, we can never reject equality across genders. Thus, the cross-country evidence suggests that grammatical gender predicts gender differences in behavior (specifically, involvement in the labor force), but also predicts traditional gender attitudes among both men and women.

5.5 Robust Inference

In this section, we discuss two potential concerns with our cross-country analysis. First, as discussed above, we were unable to classify the gender structure of some languages. Though these language tend to be small (in terms of numbers of native speakers), they account for more than one percent of the population in 22 countries. In Section 5.5.1, we present estimation that adjusts for the interval nature of our independent variable of interest, the proportion of each country’s population whose native language is a gender language. In Section 5.5.2, we consider the fact that language structures may be correlated within language families, since modern tongues evolved from common ancestors (Roberts, Winters, and Chen 2015). To address the potential correlation within families while maximizing statistical power (by exploiting variation in grammatical gender both across and between families), we introduce a permutation test based on the structure of the language tree.

5.5.1 Measurement Error

In our cross-country analysis, our independent variable of interest is the proportion of the population whose native language is a gender language. However, as discussed above, we are unable to find information on the grammatical structure of many of the world’s smaller languages. Though these unclassified languages account for less than one percent of the world population, they make up a substantial fraction of the population in a small number

of countries (e.g. Chad and Papua New Guinea). Even in countries where we successfully classify the gender structure of almost everyone, our independent variable of interest is an interval rather than a point in 85 of 193 countries — because the proportion of native speakers whose languages we classify is less than one.

This is a case described by Horowitz and Manski (1998) as “censoring of regressors,” discussed further by Aucejo, Bugni, and Hotz (2017). Our analysis so far assumes that this missingness is ignorable. Without this assumption, however, we can still estimate worst-case bounds for the maximum and minimum possible values of the parameter of interest; following Imbens and Manski (2004), we can construct a confidence interval around these bounds.

We use numerical optimization to search the space of possible independent variable values to establish worst-case upper and lower bounds, $\hat{\beta}^u$ and $\hat{\beta}^l$, that would result from estimation of Equation 10.²⁶ We then use the associated standard errors on these extrema to compute a confidence interval, employing a formula analogous to that of Equations 6 and 7 in Imbens and Manski (2004). A confidence interval with coverage probability α is equal to:

$$CI_\alpha = [\hat{\beta}^l - \bar{C} \cdot SE(\hat{\beta}^l), \hat{\beta}^u + \bar{C} \cdot SE(\hat{\beta}^u)] \quad (11)$$

where \bar{C} satisfies

$$CDF\left(\bar{C} + \frac{\hat{\Delta}}{\max(SE(\hat{\beta}^l), SE(\hat{\beta}^u))}\right) - CDF(-\bar{C}) = \alpha \quad (12)$$

for the CDF of Student’s t-distribution with the appropriate number of degrees of freedom.²⁷ Intuitively, the Manski and Imbens approach formalizes a method for shortening each end of the confidence interval relative to the union of the OLS confidence intervals around the worst-case point estimates, since the union would include the true parameter value with

²⁶We use MATLAB’s `fmincon` interior point algorithm, and confirm results using a simple hill-climbing algorithm in Stata.

²⁷Imbens and Manski do this using the normal distribution, but using the Student t-distribution yields a wider, more conservative confidence interval.

probability above 0.95 in either worst-case scenario.

In Table 6, we compare naïve OLS confidence intervals with the more conservative Imbens-Manski confidence intervals which adjust for censoring of the regressor of interest. As expected, confidence intervals widen slightly, but patterns of significance are unchanged: those confidence intervals that did not include zero in the naïve specification do not include zero after adjusting for censoring. This result is largely as expected since missing data problems are relatively minor in most countries. However, if one attempted the same bounding exercise without our data set, using only the data available in the World Atlas of Language Structures, the Imbens-Manski confidence intervals would always include zero. Thus, our data set allows for more robust inference than had previously been possible.

5.5.2 Non-Independence within Language Families

A more serious inference concern arises from the fact that languages are not independent. Different tongues evolve over time from a common ancestor. Grammatical structures vary both across and within language families. Roberts, Winters, and Chen (2015) consider a range of approaches to correcting for the non-independence of modern languages. Many approaches have the drawback that they are statistically less powerful than they could otherwise be because they ignore variation in grammatical structure either within or between language families.

We propose a permutation test approach based on the observed structure of the language tree, as documented by the *Ethnologue*. Specifically, we cluster together languages up to the highest tree level at which we observe no variation in our treatment of interest, grammatical gender. That is, we form the largest possible clusters that are homogeneous in terms of grammatical gender. Thus, for entire top-level language families that show no variation in gender structure (e.g., the Austronesian language family), we cluster at the language family level. In intermediate cases, we designate clusters at the highest level of the tree where we do not observe variation in grammatical gender (e.g., all Western Nilotic languages cluster together; they are only a branch within the Eastern Sudanic part of the Nilo-Saharan family,

which itself contains a number of other such clusters by our definition). In cases where two languages that differ in their gender structure otherwise share the same classification path through the entire language tree, we cluster at the language level.

Figure 5 illustrates this approach for a hypothetical language family. All of the languages in the Group A branch in the figure are gender languages, so they are assigned to a single cluster. Similarly, all of the languages on the Group C branch are non-gender, so they also represent a single cluster. Within Group B, the B1 languages show language-level variation: Languages B1.1 and B1.2 share the same path for the entire language tree, but they differ in gender structure. Thus, within the B1 branch of this hypothetical tree, individual languages are assigned to unique clusters. Finally, the B2 languages are all gender languages, so they are assigned to a single cluster that is distinct from the B1 clusters. Thus, the hypothetical language tree presented in the figure is partitioned into six clusters, each representing a sub-tree within the language tree that shows no gender variation.

This approach defines a set of 203 clusters, 69 of which have grammatical gender. Having assigned all the languages to clusters in this manner, we conduct a permutation test by randomly generating alternative (hypothetical) allocations of gender structure that would be possible while holding fixed the structure of the treatment variation across the language tree and the number of clusters “treated” with grammatical gender (69 of 203). We use each such hypothetical assignment of treatments to create an associated country-level measure of grammatical gender (which would be observed if treatments were assigned according to our hypothetical allocation rule, given the structure of the language tree and the distribution of languages across countries). We repeat this process 10,000 times, allowing us to estimate the likelihood that the observed associations between grammatical gender and outcomes are spurious, given the structure of the language tree, the correlation in treatment within language families, and the distribution of languages across countries.

In Table 7, we compare naïve OLS p-values to those that result from our permutation test. It is clear that appropriate clustering matters: permutation test p-values are substantially higher than the naïve OLS p-values. Nevertheless, the negative association between

grammatical gender and women’s labor force participation is still statistically significant after adjusting for the non-independence of languages. Figure 6 illustrates the full distribution of coefficient estimates under the null, highlighting the small fraction that exceed the magnitude of the true estimated coefficients. The relationships between grammatical gender and (i) the gender gap in primary school completion and (ii) gender attitudes also remain marginally significant. Thus, our results do not appear to be driven by the correlation in grammatical structure observed within language families.

6 Within-Country Analysis

6.1 Empirical Strategy

Next, we explore the relationship between gender languages and women’s labor force participation at the individual level in two contexts where both gender and non-gender languages are indigenous: sub-Saharan Africa and India. There are seven African countries where between 10 and 90 percent of the population speaks a gender native language: Chad, Kenya, Mauritania, Niger, Nigeria, South Sudan, and Uganda.²⁸ In these countries, both gender and non-gender languages are indigenous — in contrast to, for example, several countries in South America where non-gender indigenous languages and a gender colonial language are both widely spoken. The same is true in India, where 62 percent of the population speaks a gender language as their mother tongue (Lewis, Simons, and Fennig, eds., 2016). Both the Dravidian language family and the Indo-Aryan branch of the Indo-European family include both gender and non-gender languages (Masica 1991, Krishnamurti 2001). Hence, both India and sub-Saharan Africa allow us to examine the relationship between grammatical gender and women’s outcomes while holding a much of the cultural and institutional context constant.

We use two data sources in our within-country analysis: the Afrobarometer surveys (Afrobarometer Data 2016) and the India Human Development Survey (Desai, Dubey, and

²⁸To provide an example of the variation within the Nilo-Saharan language family: Maasai is a gender language; Luo, however, is not.

Vanneman 2015). Of the seven African countries listed above, we focus on the four that have been included in at least one round of the Afrobarometer survey: Kenya, Niger, Nigeria, and Uganda. Four rounds of data are available for Kenya, Nigeria, and Uganda, while only one round of data is available for Niger.²⁹ Our sample includes 26,546 Afrobarometer respondents who speak 175 different languages. Our IHDS sample includes 75,966 household heads and their spouses living in 33 Indian states. IHDS respondents in our sample speak 61 distinct Indian languages.

Our individual-level analysis parallels our cross-country analysis. We consider two main outcomes: labor force participation (an indicator equal to one if a respondent either does some type of income-generating activity or is actively looking for a job) and education (indicators for having completed primary and secondary school). We report two regression specifications. First, we estimate the association between grammatical gender and each outcome of interest in a sample of (only) women, estimating the OLS regression equation:

$$Y_i = \alpha + \beta \text{Gender}_i + \gamma Z_i + \varepsilon_i \quad (13)$$

where Y_i is the outcome of interest for woman i , Gender_i is an indicator for having a gender language as one's mother tongue, Z_i is a vector of controls (age, age², and a set of religion dummies), and ε_i is a mean-zero error term. In our analysis of the Afrobarometer data, we also include a country-by-survey-round fixed effects. As in our cross-country analysis, we wish to avoid confounding the impact of grammatical gender on women's education and labor force participation with other cultural factors that might impact both outcomes for both men and women. To do this, we also report pooled OLS regressions that include data on both men and women. These take the form:

$$Y_i = \alpha + \beta \text{Gender}_i + \zeta \text{Female}_i + \mu \text{Gender} \times \text{Female}_i + \gamma Z_i + \varepsilon_i \quad (14)$$

²⁹The first round of the Afrobarometer surveys did not include sufficiently detailed data on native languages for inclusion in our analysis. Our analysis includes data from Afrobarometer Rounds 2 through 5 for Kenya, Nigeria, and Uganda. Niger was only added to the Afrobarometer in Round 5; that round is included in our analysis.

where $Gender \times Female_i$ is an interaction between a female dummy and the indicator for being a native speaker of a gender language. In these specifications, we also include interactions between the $Female_i$ dummy and our age and religion controls. Throughout our analysis, we cluster standard errors by language.

6.2 Results

We summarize our regression results in Figure 7 (regression results are presented in Online Appendix Tables A5 through A12). Panel A presents results on women’s labor force participation. In the Afrobarometer data, we see a negative and statistically significant relationship between grammatical gender and both levels of and gender differences in labor force participation. Coefficient estimates are broadly similar in the Indian data, particularly the estimates of gender differences in labor force participation. However, the relationship is not statistically significant after clustering at the language level. Turning to primary school completion (Panel B of Figure 7), we see that grammatical gender is negatively and significantly related to both rates of primary school completion and the gender difference in primary school completion in both Sub-Saharan Africa and India. Coefficient estimates suggest that having a gender mother tongue is associated with more than a 10 percentage point decline in the likelihood that a woman completed primary school. We see a more muted association between grammatical gender and secondary school completion (Panel C of Figure 7), though results still suggest a negative and statistically significant relationship in both the African and the Indian data — particularly after controlling for completion rates among men within the same ethnolinguistic group. Thus, in both Africa and India, we see that the cross-country pattern is largely replicated within country, even when restricting attention to indigenous languages that differ in terms of their grammatical gender structure.

7 Causality

The analysis presented thus far documents a strongly negative cross-country relationship between grammatical gender and women’s labor force participation, and shows that it is robust to a permutation test that addresses the potential non-independence of languages. We also document a positive cross-country relationship between grammatical gender and traditional gender attitudes, and a marginally statistically significant cross-country relationship between grammatical gender and the gender gap in primary school completion. We then find that the negative associations between grammatical gender and women’s educational attainment and labor force participation are replicated within four African countries and within India. The caveat, of course, is that all of these are correlations, and not necessarily causal relationships.

In most cases, whether a language has retained grammatical gender is driven by idiosyncracies of history far-removed from outcomes of interest in this paper. For example, scholars believe that English lost grammatical gender because its complex declensional agreement system eroded over time, in part because of the influx of Scandinavian immigrants (who learned English as a second language in adulthood) into the linguistic community (McWhorter 2005, Kastovsky 1999). So, English did not lose grammatical gender because of changes in gender norms in pre-Norman England. Nevertheless, gender languages are not randomly assigned. The observed correlations may be driven by some unobserved causal factor that is correlated with both language and gender norms.

To assess whether the observed correlation is likely to represent a causal link between language and our outcomes of interest, we follow the approach suggested by Altonji, Elder, and Taber (2005) and further refined by Oster (2017).³⁰ Under the assumption that the relationship between the outcome variables, treatment, and the observed controls is similar to the relationship between the outcomes, treatment, and unobserved controls, this approach relates changes in coefficient magnitudes as controls are added to changes in the

³⁰An alternative approach would be to try and identify a suitable instrument for grammatical gender. However, recent work suggests that conventional approaches may overstate the precision of 2SLS estimates, leading to invalid inference (Young 2018). Thus, OLS with caution may be an equally reasonable approach.

observed R^2 . Intuitively, omitted variable bias is assumed to be proportional to changes in regression coefficients as controls are added; however, these changes must be scaled by changes in the R^2 — adding controls that do not explain the outcome variable does little to address concerns about omitted variable bias.

Following the procedures outlined by Oster (2017), we estimate two measures of coefficient stability. These additional statistics are calculated using the results from two OLS regressions: (i) a bivariate regression of an outcome of interest on grammatical gender, which generates a coefficient of interest, $\hat{\beta}$, and an associated \hat{R}^2 ; and (ii) a multivariate regression of the same outcome on grammatical gender plus a set of controls, which generates a second OLS coefficient, $\tilde{\beta}$, and an associated \tilde{R}^2 .

In this framework, δ^* is the proportional selection coefficient. Given the empirical relationship between the outcome, the treatment, and the observed controls, δ^* indicates how much more correlated with treatment the unobservables would need to be in order to explain the entire association between treatment and the outcome of interest. If $\delta^* > 1$, then an observed empirical relationship is relatively robust in that unobservables would need to be more correlated with treatment than observables to explain the association. A second parameter of interest is β^* . It indicates the likely causal impact of grammatical gender on an outcome of interest under the assumption that $\delta^* = 1$ (i.e. assuming that the covariance structure is the same for observables and unobservables).

Coefficient stability results are presented in Table 8. Cross-country results are presented in Panel A. Results indicate that our estimates of the impact of grammatical gender on women’s labor force participation are unlikely to be driven by selection alone. Unobservables would need to be 1.44 times more correlated with treatment (than observables) to explain the observed link between grammatical gender and the level of women’s labor force participation; unobservables would need to be 3.24 times more correlated with treatment to explain the gender gap in labor force participation. Thus, the analysis suggests that gender languages reduce women’s labor force participation in both absolute and relative terms.

Our individual-level analysis is presented in Panel B (Afrobarometer data) and Panel

C (IHDS data for India) of Table 8. In all cases, the Oster (2017) approach suggests that the empirical relationship between grammatical gender and outcomes of interest is unlikely to be driven by selection on unobservables.

Thus, the coefficient stability approach supports the hypothesis that grammatical gender has a causal impact on women’s labor force participation and, in India and parts of Sub-Saharan Africa, women’s educational attainment. Nevertheless, this approach — like instrumental variables — relies on fundamentally untestable assumptions. Though modern gender attitudes could not plausibly have impacted the grammatical structure of language, we cannot fully rule out the possibility that cultural factors shaped both grammatical structure and gender norms. As in all studies of history and culture, it is not possible to run experiments and relevant sample sizes are fairly small; some measure of caution about causal claims is therefore warranted.

8 Conclusion

Using a new data set on the grammatical gender structure of more than 4,000 languages, we document a robust negative association between gender languages and women’s labor force participation. At the country level, an increase in the proportion of the population whose native language is a gender language is associated with lower female labor force participation and — perhaps more importantly — larger gender differences in labor force participation. Using data from the World Values Survey, we show that grammatical gender also predicts support for traditional gender roles. The prevalence of gender languages is also related to gender gaps in primary school completion, though the association is only marginally statistically significant.

Focusing on five countries where both gender and non-gender languages are indigenous and widely spoken (India, Kenya, Niger, Nigeria, and Uganda), we show that a similar pattern holds within countries. Speaking a gender native language is associated with lower labor force participation and educational attainment among women, both in absolute terms

and relative to men from the same ethnolinguistic group. Both our cross-country and our individual-level regressions are robust to the inclusion of controls that could not plausibly have been impacted by treatment; if one is willing to assume that the relationship between unobserved omitted factors, treatment, and the outcomes of interest is similar to the observed relationship between controls, treatment, and the outcomes of interest, our estimates suggest that grammatical gender has a large negative impact on women’s labor force participation.

Our results are consistent with research in psychology, linguistics, and anthropology suggesting that languages shape patterns of thought in subtle and subconscious ways. Languages are a critical part of our cultural heritage, and it would be inappropriate to suggest that some languages are detrimental to development or women’s rights. However, languages do evolve over time; and the direction of their evolution is shaped by both individual choices (for example, whether to use gendered pronouns like “he” or “she” or gender-neutral alternatives such as “they”) and conscious decisions by government agencies (e.g. the Académie Française) and other thought leaders (e.g. major newspapers and magazines). Our results suggest that individuals should reflect upon the social consequences of their linguistic choices, as the nature of the language we speak shapes the ways we think, and the ways our children will think in the future.

References

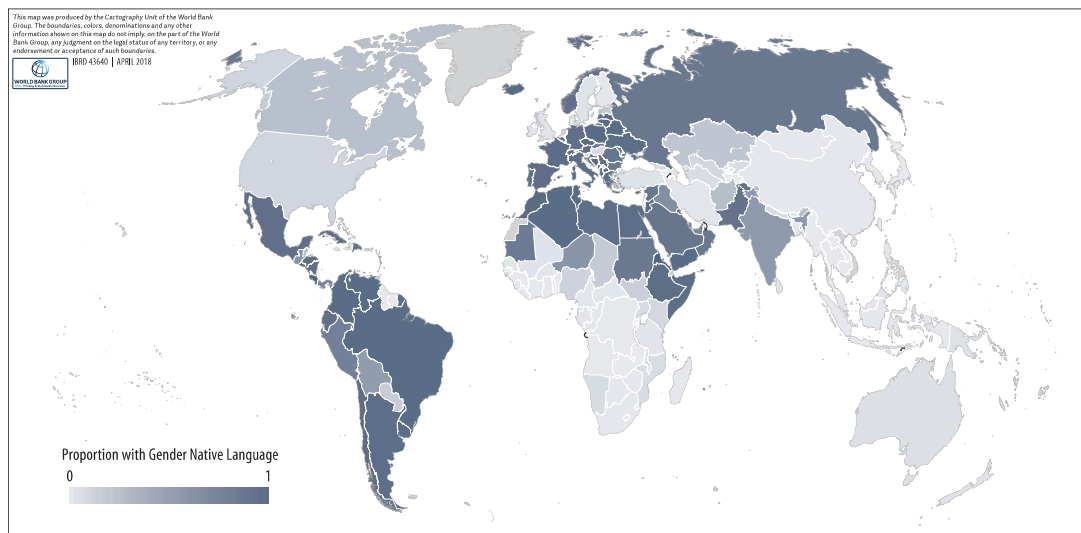
- ACHARYA, A., M. BLACKWELL, AND M. SEN (2016): “Explaining Causal Findings without Bias: Detecting and Assessing Direct Effects,” *American Political Science Review*, 110(3), 512–529.
- AFROBAROMETER DATA (2016): “Kenya, Niger, Nigeria, Tanzania, Uganda, Rounds 2 through 5, 2002–2013,” available at <http://www.afrobarometer.org>.
- AIKHENVALD, A. Y. (2003): *Classifiers: A Typology of Noun Categorization Devices*. Oxford University Press, Oxford, UK.
- ALESINA, A., P. GIULIANO, AND N. NUNN (2013): “On the Origins of Gender Roles: Women and the Plough,” *Quarterly Journal of Economics*, 128(2), 469–530.
- ALTONJI, J. G., T. E. ELDER, AND C. R. TABER (2005): “Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools,” *Journal of Political Economy*, 113(1), 151–184.
- ANGRIST, J. D., AND J.-S. PISCHKE (2008): *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton University Press.
- AUCEJO, E. M., F. A. BUGNI, AND V. J. HOTZ (2017): “Identification and inference on regressions with missing covariate data,” *Econometric Theory*, 33(1), 196–241.
- BARRO, R., AND J.-W. LEE (2013): “A New Data Set of Educational Attainment in the World, 1950–2010,” *Journal of Development Economics*, 104, 184–198.
- BEAMES, J. (1875): *Comparative Grammar of the Modern Aryan Languages of India: to wit, Hindi, Punjabi, Sindhi, Gujarati, Marathi, Oriya, and Bangali: Vol. II. The Noun and Pronoun*. Trubner and Company, London, U.K.
- BORODITSKY, L., L. A. SCHMIDT, AND W. PHILLIPS (2003): “Sex, Syntax, and Semantics,” in *Language in Mind: Advances in the Study of Language and Thought*, ed. by S. Goldin-Meadow, and D. Gentner, pp. 61–79. MIT Press, Cambridge, MA.
- CAMPBELL, G. L. (1991): *Compendium of the World’s Languages*. Routledge, London, UK.
- CHEN, M. K. (2013): “The Effect of Language on Economic Behavior: Evidence from Savings Rates, Health Behaviors, and Retirement Assets,” *American Economic Review*, 103(2), 690–731.
- CORBETT, G. G. (1991): *Gender*. Cambridge University Press, Cambridge, UK.
- (2013a): “Number of Genders,” in *The World Atlas of Language Structures Online*, ed. by M. S. Dryer, and M. Haspelmath. Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.
- (2013b): “Sex-based and Non-sex-based Gender Systems,” in *The World Atlas of Language Structures Online*, ed. by M. S. Dryer, and M. Haspelmath. Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.
- (2013c): “Systems of Gender Assignment,” in *The World Atlas of Language Structures Online*, ed. by M. S. Dryer, and M. Haspelmath. Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.
- CREISSELS, D. (2000): “Typology,” in *African Languages: An Introduction*, ed. by B. Heine, and D. Nurse, pp. 231–258. Cambridge University Press, Cambridge, UK.
- DANZIGER, S., AND R. WARD (2010): “Language Changes Implicit Associations between Ethnic Groups and Evaluation in Bilinguals,” *Psychological Science*, 21(6), 799–800.

- DESAI, S., A. DUBEY, AND R. VANNEMAN (2015): “India Human Development Survey-II (IHDS-II),” University of Maryland and National Council of Applied Economic Research, New Delhi, distributed by: Inter-university Consortium for Political and Social Research, Ann Arbor, MI.
- DURKHEIM, E., AND M. MAUSS (1963): *Primitive Classification*. University of Chicago Press, Chicago, USA, Translated from the French, edited, and with an introduction by Rodney Needham.
- FRYER, R. G., AND S. D. LEVITT (2010): “An Empirical Analysis of the Gender Gap in Mathematics,” *American Economic Journal: Applied Economics*, 2(2), 210–40.
- GALOR, O., O. OZAK, AND A. SARID (2018): “Geographical Roots of the Coevolution of Cultural and Linguistic Traits,” Working Paper 25289, National Bureau of Economic Research.
- GAY, V., D. L. HICKS, E. SANTACREU-VASUT, AND A. SHOHAM (2017): “Decomposing culture: An Analysis of Gender, Language, and Labor Supply in the Household,” *Review of Economics of the Household*, 16(4), 879–909.
- GIVATI, Y., AND U. TROIANO (2012): “Law, Economics, and Culture: Theory of Mandated Benefits and Evidence from Maternity Leave Policies,” *Journal of Law and Economics*, 55(2), 339–364.
- GRANT, M. J., AND J. R. BEHRMAN (2010): “Gender Gaps in Educational Attainment in Less Developed Countries,” *Population and Development Review*, 36(1), 71–89.
- GRIERSON, G. A. (1903a): *Linguistic Survey of India: Volume V, Indo-Aryan Family Eastern Group, Part I, Specimens of the Bengali and Assamese Languages*. Superintendent Government Printing, Calcutta, India, <http://dsal.uchicago.edu/books/lsi/index.html> [accessed 13 July 2016].
- (1903b): *Linguistic Survey of India: Volume V, Indo-Aryan Family Eastern Group, Part II, Specimens of the Bihari and Oriya Languages*. Superintendent Government Printing, Calcutta, India, <http://dsal.uchicago.edu/books/lsi/index.html> [accessed 13 July 2016].
- (1904): *Linguistic Survey of India: Volume VI, Indo-Aryan Family Mediate Group, Specimens of the Eastern Hindi Language*. Superintendent Government Printing, Calcutta, India, <http://dsal.uchicago.edu/books/lsi/index.html> [accessed 13 July 2016].
- (1905): *Linguistic Survey of India: Volume VII, Indo-Aryan Family Southern Group, Specimens of the Marathi Language*. Superintendent Government Printing, Calcutta, India, <http://dsal.uchicago.edu/books/lsi/index.html> [accessed 13 July 2016].
- (1907): *Linguistic Survey of India: Volume IX, Indo-Aryan Family Central Group, Part III, The Bhil Languages*. Superintendent Government Printing, Calcutta, India, <http://dsal.uchicago.edu/books/lsi/index.html> [accessed 13 July 2016].
- (1908): *Linguistic Survey of India: Volume IX, Indo-Aryan Family Central Group, Part II, Specimens of Rajastani and Gujarati*. Superintendent Government Printing, Calcutta, India, <http://dsal.uchicago.edu/books/lsi/index.html> [accessed 13 July 2016].
- (1909): *Linguistic Survey of India: Volume III, Tibeto-Burman Family, Part I, General Introduction, Specimens of the Tibetan Dialects, the Himalayan Dialects, and the North Assam Group*. Superintendent Government Printing, Calcutta, India, <http://dsal.uchicago.edu/books/lsi/index.html> [accessed 13 July 2016].
- (1916): *Linguistic Survey of India: Volume IX, Indo-Aryan Family Central Group, Part I, Specimens of Western Hindi and Panjabi*. Superintendent Government Printing, Calcutta, India, <http://dsal.uchicago.edu/books/lsi/index.html> [accessed 13 July 2016].

- (1919): *Linguistic Survey of India: Volume VIII, Indo-Aryan Family Northwestern Group, Part II, Specimens of the Dardic or Pisacha Languages*. Superintendent Government Printing, Calcutta, India, <http://dsal.uchicago.edu/books/lsi/index.html> [accessed 13 July 2016].
- (1921): *Linguistic Survey of India: Volume X, Specimens of Languages of the Eranian Family*. Superintendent Government Printing, Calcutta, India, <http://dsal.uchicago.edu/books/lsi/index.html> [accessed 13 July 2016].
- GUIORA, A. Z. (1983): “Language and Concept Formation: A Cross-Lingual Analysis,” *Cross-Cultural Research*, 18(3), 228–256.
- HELLINGER, M. (2003): “English,” in *Gender Across Languages: The Linguistic Representation of Women and Men, Volume*, ed. by M. Hellinger, and H. Bußman, pp. 105–114. John Benjamins Publishing Company, Amsterdam, The Netherlands.
- HELLINGER, M., AND H. BUSSMAN (2003): “The Linguistic Representation of Women and Men,” in *Gender Across Languages: The Linguistic Representation of Women and Men, Volume 3*, ed. by M. Hellinger, and H. Bußman, pp. 1–26. John Benjamins Publishing Company, Amsterdam, The Netherlands.
- HICKS, D. L., E. SANTACREU-VASUT, AND A. SHOHAM (2015): “Does Mother Tongue Make for Women’s Work? Linguistics, Household Labor, and Gender Identity,” *Journal of Economic Behavior and Organization*, 110(2), 19–44.
- HOROWITZ, J. L., AND C. F. MANSKI (1998): “Censoring of outcomes and regressors due to survey nonresponse: Identification and estimation using weights and imputations,” *Journal of Econometrics*, 84(1), 37–58.
- IMBENS, G. W., AND C. F. MANSKI (2004): “Confidence Intervals for Partially Identified Parameters,” *Journal of Econometrics*, 72(6), 1845–1857.
- JANHUNEN, J. (1999): “Grammatical Gender from East to West,” in *Gender in Grammar and Cognition*, ed. by B. Unterbeck, and M. Rissanen, pp. 689–708. Mouton de Gruyter, Berlin, Germany.
- KASTOVSKY, D. (1999): “Inflectional Classes, Morphological Restructuring, and the Dissolution of Old English Grammatical Gender,” in *Gender in Grammar and Cognition*, ed. by B. Unterbeck, and M. Rissanen, pp. 709–728. Mouton de Gruyter, Berlin, Germany.
- KILARSKI, M. (2013): *Nominal Classification: A History of Its Study from the Classical Period to the Present*. John Benjamins Publishing Company, Amsterdam, The Netherlands.
- KRISHNAMURTI, B. (2001): *Comparative Dravidian Linguistics*. Oxford University Press, Oxford, UK.
- LAKOFF, G. (1987): *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. University of Chicago Press, Chicago, IL.
- LANCY, D. F. (2015): *The Anthropology of Childhood: Cherubs, Chattle, Changelings*. Cambridge University Press, Cambridge, UK.
- LEWIS, M. P., G. F. SIMONS, AND C. D. FENNIG, EDS., (2016): “Ethnologue: Languages of the World, nineteenth edition.” Dallas, Texas: SIL International, Online version: <http://www.ethnologue.com> [accessed 4 June 2016].
- MASICA, C. P. (1991): *The Indo-Aryan Languages*. Cambridge University Press, Cambridge, UK.
- MAVISAKALYAN, A. (2015): “Gender in Language and Gender in Employment,” *Oxford Development Studies*, 43(4), 403–424.

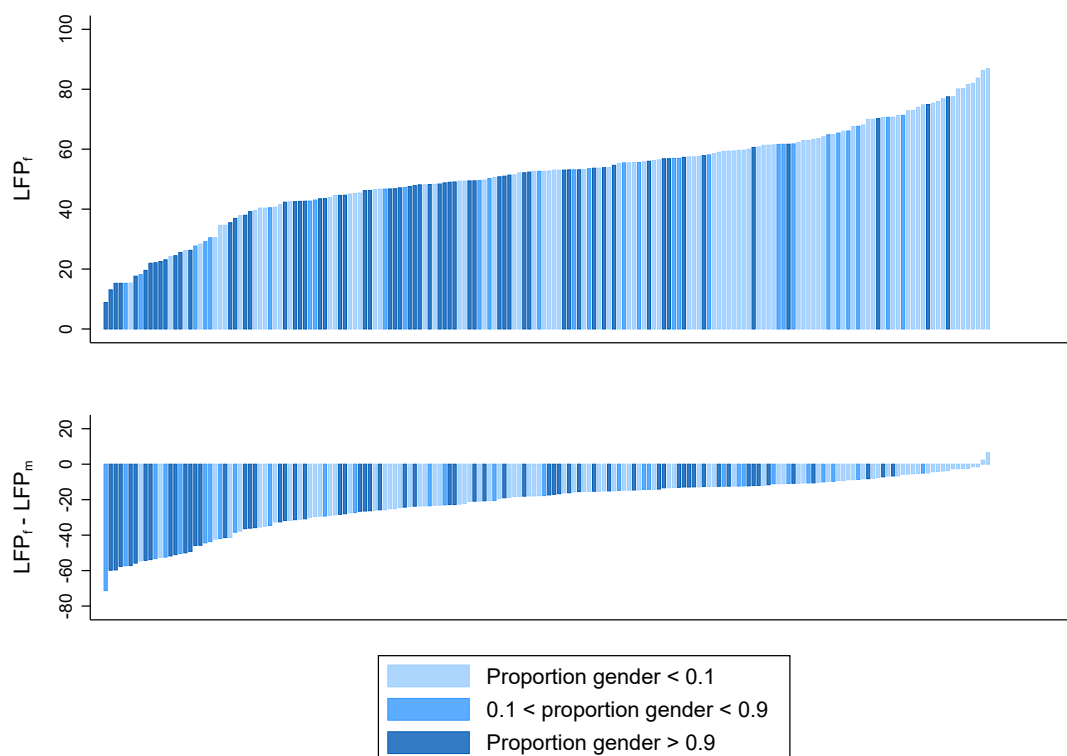
- MCWHORTER, J. H. (2005): *Defining Creole*. Oxford University Press.
- MUHLEISEN, S., AND D. E. WALICEK (2010): “Language and Gender in the Caribbean: An Overview,” *Sargasso: Explorations of Language, Gender, and Sexuality*, 2008–2009(I), 15–30.
- OGUNNAIKE, O., Y. DUNHAM, AND M. R. BANAJI (2010): “The Language of Implicit Preferences,” *Journal of Experimental Social Psychology*, 46(6), 999–1003.
- OSTER, E. (2017): “Unobservable Selection and Coefficient Stability: Theory and Evidence,” *Journal of Business and Economic Statistics*.
- PÉREZ, E. O., AND M. TAVITS (forthcoming): “Language Influences Public Attitudes Toward Gender Equality,” *Journal of Politics*.
- ROBERTS, S. G., J. WINTERS, AND K. CHEN (2015): “Future Tense and Economic Decisions: Controlling for Cultural Evolution,” *PLOS One*, 10(7), e0132145.
- SANTACREU-VASUT, E., O. SHENKAR, AND A. SHOHAM (2014): “Linguistic Gender Marking and Its International Business Ramifications,” *Journal of International Business Studies*, 45, 1170–1178.
- SANTACREU-VASUT, E., A. SHOHAM, AND V. GAY (2013): “Do female/male distinctions in language matter? Evidence from gender political quotas,” *Applied Economics Letters*, 20(5), 495–498.
- SHOHAM, A., AND S. M. LEE (2017): “The Causal Impact of Grammatical Gender Marking on Gender Wage Inequality and Country Income Inequality,” *Business and Society*.
- WHORF, B. L. (2011[1956]a): “Language, Mind, and Reality,” in *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, ed. by J. B. Carroll, pp. 246–270. Martino Publishing, Mansfield Centre, CT.
- (2011[1956]b): “A Linguistic Consideration of Thinking in Primitive Communities,” in *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, ed. by J. B. Carroll, pp. 65–86. Martino Publishing, Mansfield Centre, CT.
- WORLD VALUES SURVEY ASSOCIATION (2015): “World Values Survey Wave 6 2010–2014 Official Aggregate v.20150418,” available online at www.worldvaluessurvey.org, aggregate file producer: Asep/JDS, Madrid SPAIN.
- YOUNG, A. (2018): “Consistency without Inference: Instrumental Variables in Practical Application,” working paper.

Figure 2: The Distribution of Gender Languages



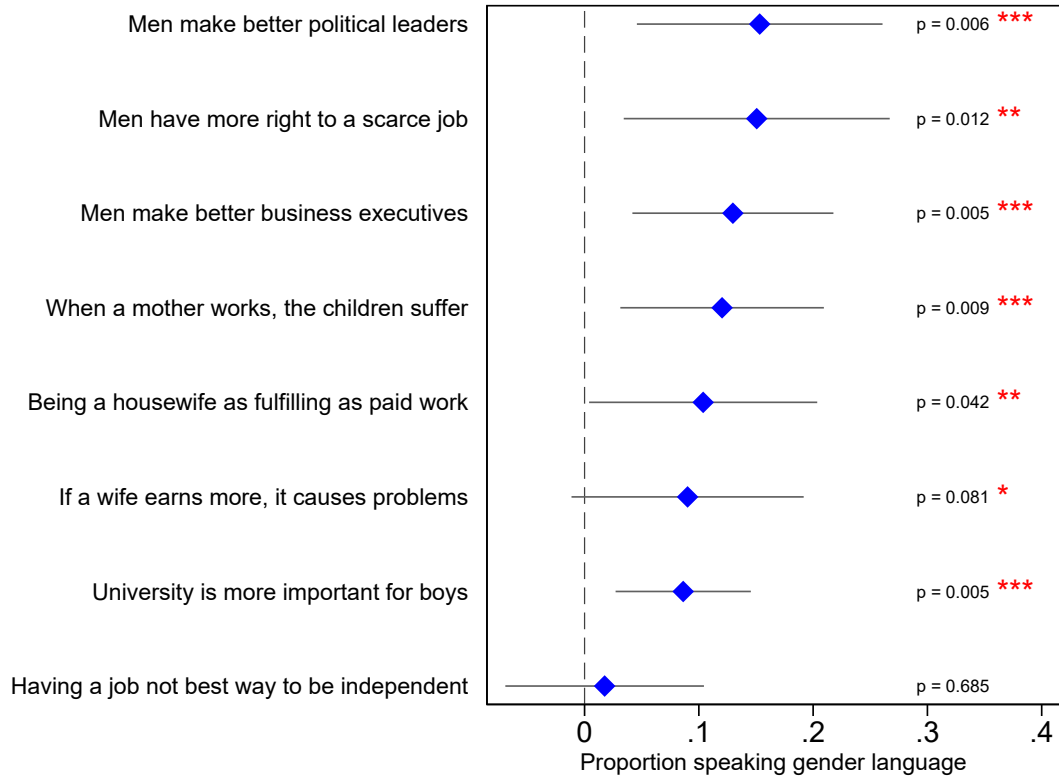
The figure shows the percentage of the native speakers in each country whose native language is a gender language (i.e. the fraction of Ethnologue native speakers whose native language uses a system of grammatical gender). The figure assumes that missing data (on 0.8 percent of all native speakers worldwide) is ignorable.

Figure 3: Cross-Country Variation in Female Labor Force Participation



The figure plots the level of female labor force participation (top panel) and the gender difference in labor force participation (bottom panel) by country. Darker bars indicate countries with a higher proportion of native speakers of gender languages.

Figure 4: Cross-Country Variation in Gender Attitudes



The figure summarizes the results from a series of regressions of (country-level averages of) responses to World Values Survey (WVS) questions on the proportion of a country's population whose native language is a gender language. We present the results for all eight WVS questions related to gender attitudes. Responses to all eight questions are coded so that the answer most consistent with traditional gender norms (involving separate roles for men and women) is equal to 1 and the response most consistent with gender equality is equal to 0. Each regression is estimated via OLS and includes continent fixed effects. The outcome in the first row is the average response to the question "When a mother works for pay, the children suffer" (agreement is coded as a 1, disagreement as a 0). The outcome variable in the second row is the average response to the statement "When jobs are scarce, men should have more right to a job than women." In the third row, the outcome variable is based on the statement "On the whole, men make better political leaders than women do." In the fourth row, the outcome variable is based on the statement "On the whole, men make better business executives than women do." In the fifth row, the outcome variable is based on the statement "Being a housewife is just as fulfilling as working for pay;" agreement was coded as 0 and disagreement was coded as 1. In the sixth row, the outcome variable is based on the statement "If a woman earns more money than her husband, it's almost certain to cause problems." In the seventh row, the outcome variable is based on the statement "A university education is more important for a boy than for a girl." In the last row, the outcome variable is based on the statement "Having a job is the best way for a woman to be an independent person;" in this case, disagreement was coded as 1 and agreement was coded as 0.

Figure 5: Assignment to Clusters for the Permutation Test

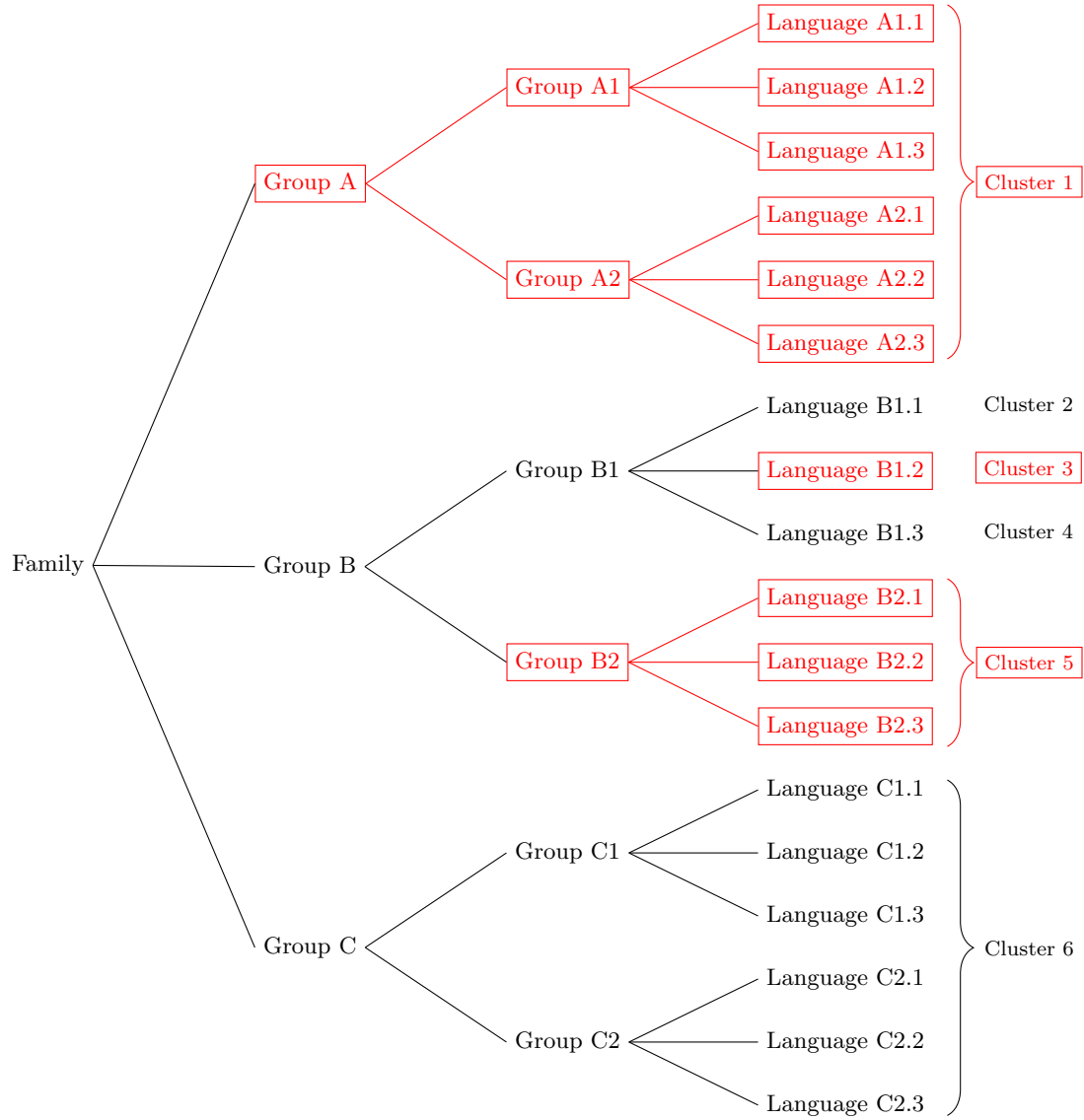
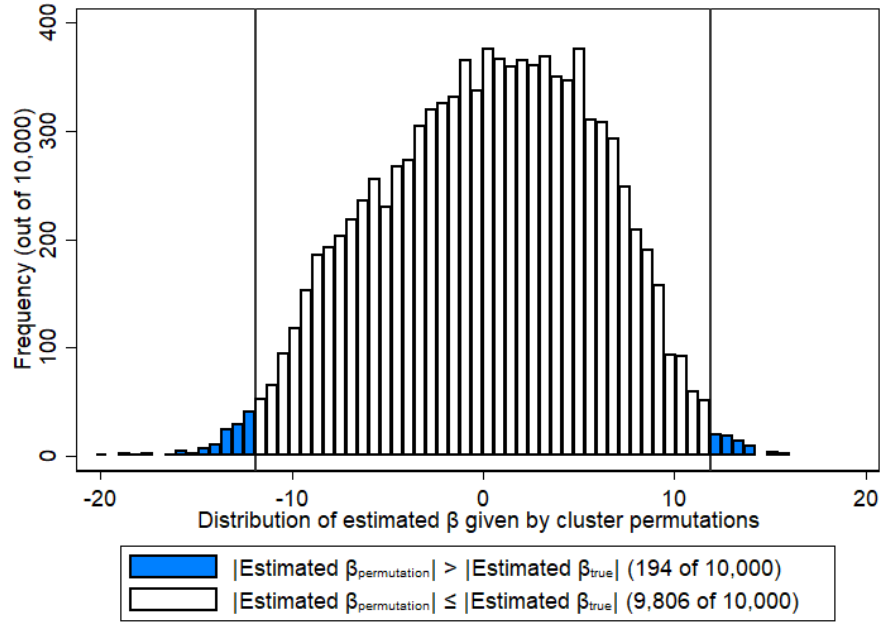


Figure illustrates a hypothetical language family. Gender languages and branches of the tree that include only gender languages are boxed and printed in red. Languages are assigned to clusters at the highest level of the language tree that shows no variation in grammatical gender.

Figure 6: Permutation Tests

Panel A: Female Labor Force Participation



Panel B: Gender Difference in Labor Force Participation

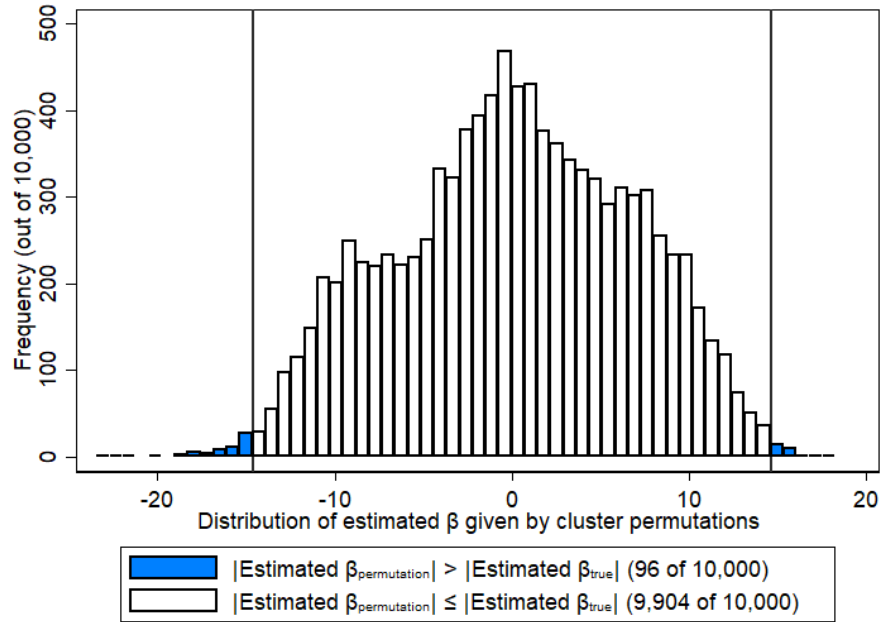
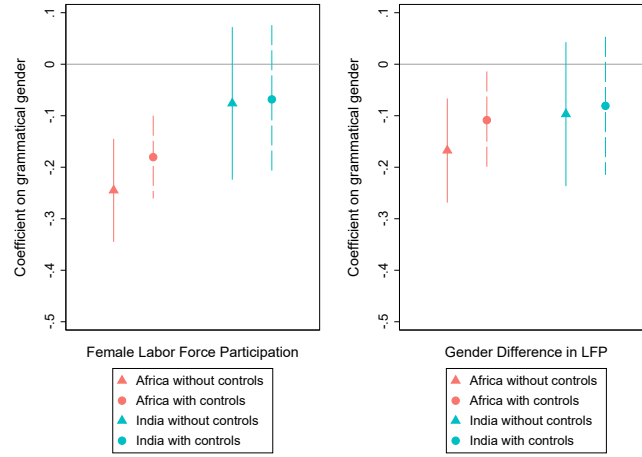
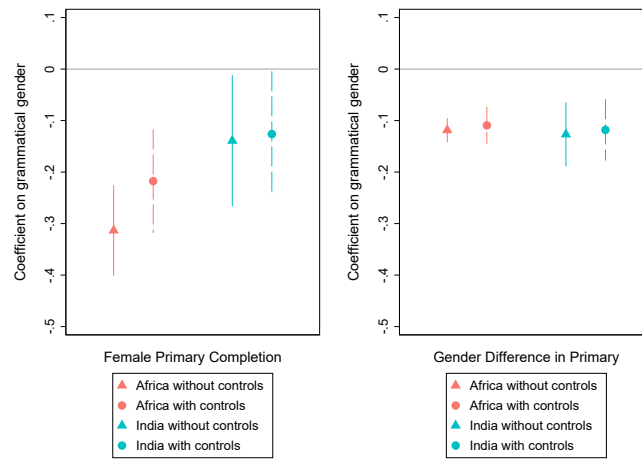


Figure 7: Within-Country Variation in Grammatical Gender

Panel A: Labor Force Participation



Panel B: Primary School Completion



Panel C: Secondary School Completion

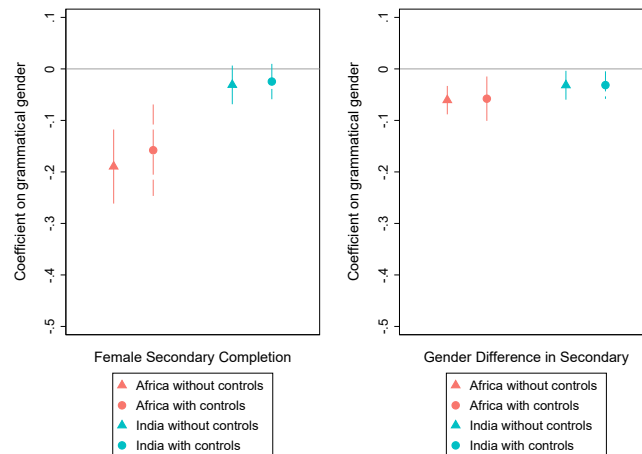


Table 1: Cross-Country OLS Regressions of Labor Force Participation

<i>Dependent variable:</i> <i>Specification:</i>	LFP _f			LFP _f - LFP _m		
	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)	OLS (6)
Proportion speaking gender language	-13.83 (2.80) [$p < 0.001$]	-17.67 (3.52) [$p < 0.001$]	-11.92 (3.34) [$p < 0.001$]	-11.61 (2.47) [$p < 0.001$]	-18.87 (3.14) [$p < 0.001$]	-14.66 (3.25) [$p < 0.001$]
Continent Fixed Effects	No	Yes	Yes	No	Yes	Yes
Country-Level Geography Controls	No	No	Yes	No	No	Yes
Observations	178	178	178	178	178	178
R^2	0.15	0.25	0.33	0.12	0.41	0.47

Robust standard errors are clustered by the most widely spoken language in all specifications; they are reported in parentheses. P-values are reported in square brackets. LFP_f is the percentage of women in the labor force, measured in 2011. LFP_f - LFP_m is the gender difference in labor force participation — i.e. the difference between female and male labor force participation, again measured in 2011. Geography controls are the percentage of land area in the tropics or subtropics, average yearly precipitation, average temperature, an indicator for being landlocked, and the Alesina *et al.* (2013) measure of suitability for the plough.

Table 2: Cross-Country OLS Regressions of Primary School Completion

<i>Dependent variable:</i> <i>Specification:</i>	PRI _f			PRI _f - PRI _m		
	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)	OLS (6)
Proportion speaking gender language	14.79 (5.83) [0.013]	-4.72 (4.44) [0.290]	-6.71 (4.40) [0.130]	1.21 (2.14) [0.573]	-3.87 (2.04) [0.060]	-3.72 (2.16) [0.088]
Continent Fixed Effects	No	Yes	Yes	No	Yes	Yes
Country-Level Geography Controls	No	No	Yes	No	No	Yes
Observations	142	142	142	142	142	142
R ²	0.06	0.53	0.61	0.00	0.18	0.20

Robust standard errors are clustered by the most widely spoken language in all specifications; they are reported in parentheses. P-values are reported in square brackets. PRI_f is the rate of primary school completion among adult women. PRI_f - PRI_m is the gender difference in primary school completion. Geography controls are the percentage of land area in the tropics or subtropics, average yearly precipitation, average temperature, an indicator for being landlocked, and the Alesina *et al.* (2013) measure of suitability for the plough.

Table 3: Cross-Country OLS Regressions of Secondary School Completion

<i>Dependent variable:</i> <i>Specification:</i>	SEC _f			SEC _f - SEC _m		
	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)	OLS (6)
Proportion speaking gender language	14.52 (5.77) [0.013]	-1.63 (4.22) [0.699]	0.43 (3.70) [0.907]	0.48 (1.93) [0.802]	0.72 (2.31) [0.756]	-0.86 (2.35) [0.716]
Continent Fixed Effects	No	Yes	Yes	No	Yes	Yes
Country-Level Geography Controls	No	No	Yes	No	No	Yes
Observations	142	142	142	142	142	142
R ²	0.06	0.47	0.67	0.00	0.07	0.10

Robust standard errors are clustered by the most widely spoken language in all specifications; they are reported in parentheses. P-values are reported in square brackets. SEC_f is the rate of secondary school completion among adult women. SEC_f - SEC_m is the gender difference in secondary school completion. Geography controls are the percentage of land area in the tropics or subtropics, average yearly precipitation, average temperature, an indicator for being landlocked, and the Alesina *et al.* (2013) measure of suitability for the plough.

Table 4: Cross-Country OLS Regressions of Gender Attitudes

<i>Dependent variable:</i>	GENDER ATTITUDES INDEX		
<i>Specification:</i>	OLS (1)	OLS (2)	OLS (3)
Proportion speaking gender language	-0.03 (0.05) [0.576]	-0.11 (0.03) [$p < 0.001$]	-0.12 (0.04) [0.002]
Continent Fixed Effects	No	Yes	Yes
Country-Level Geography Controls	No	No	Yes
Observations	56	56	56
R^2	0.01	0.74	0.78

Robust standard errors clustered by most widely spoken language in all specifications. The GENDER ATTITUDES INDEX is constructed by taking the first principal component of the 8 World Values Survey questions relating to gender norms (described in Figure 4) at the individual level, and then calculating the average of this index within a country. Numbers closer to 1 indicate more support for gender equality. Geography controls are the percentage of land area in the tropics or subtropics, average yearly precipitation, average temperature, an indicator for being landlocked, and the Alesina *et al.* (2013) measure of suitability for the plough..

Table 5: OLS Regressions of Gender Attitudes Index — Women vs. Men

<i>Dependent variable:</i>	ATTITUDES AMONG WOMEN			ATTITUDES AMONG MEN		
<i>Specification:</i>	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)	OLS (6)
Proportion speaking gender language	-0.02 (0.05) [0.714]	-0.09 (0.03) [0.002]	-0.10 (0.04) [0.012]	-0.04 (0.06) [0.508]	-0.12 (0.03) [$p < 0.001$]	-0.14 (0.04) [$p < 0.001$]
Continent Fixed Effects	No	Yes	Yes	No	Yes	Yes
Country-Level Geography Controls	No	No	Yes	No	No	Yes
Observations	56	56	56	56	56	56
R^2	0.00	0.69	0.73	0.02	0.74	0.78

Robust standard errors are clustered by the most widely spoken language in all specifications; they are reported in parentheses. P-values are reported in square brackets. The the dependent variable is constructed by taking the first principal component of the 8 World Values Survey questions relating to gender norms (described in Figure 4) at the individual level, and then calculating the average of this index by gender (i.e. separately among men and women) within a country. Geography controls are the percentage of land area in the tropics or subtropics, average yearly precipitation, average temperature, an indicator for being landlocked, and the Alesina *et al.* (2013) measure of suitability for the plough.

Table 6: Robust Inference: Manski-Imbens Worst-Case 95-Percent Confidence Intervals

	NAÏVE OLS CI	IMBENS-MANSKI CI
Female labor force participation	[−18.533, −5.305]	[−18.467, −5.013]
Gender difference in labor force participation	[−21.077, −8.233]	[−20.916, −7.741]
Female primary school completion	[−15.431, 2.010]	[−16.221, 1.673]
Gender difference in primary school completion	[−8.003, 0.559]	[−8.446, 0.432]
Female secondary school completion	[−6.901, 7.769]	[−8.261, 7.327]
Gender difference in secondary school completion	[−5.510, 3.799]	[−5.401, 3.746]
Gender attitudes index	[−0.193, −0.045]	[−0.194, −0.047]
Gender attitudes index among women	[−0.173, −0.022]	[−0.173, −0.023]
Gender attitudes index among men	[−0.214, −0.063]	[−0.215, −0.064]

Confidence intervals estimated following procedures outlined in Section 5.5.1. For each outcome, the naïve confidence interval comes from the associated regression in a previous table. The Imbens-Manski worst-case confidence interval is calculated by finding the minimum and maximum possible point estimates of the relevant coefficient based on the interval nature of the dataset (without complete data on the grammatical structure of all languages, the right-hand-side variable—the fraction of a country’s population speaking a gender language—is only observed up to an interval in some cases), then by tightening the confidence interval for correct coverage following Imbens and Manski (2004).

Table 7: Robust inference: Language structure

	NAÏVE OLS P-VALUES	PERMUTATION-BASED P-VALUES
Female labor force participation	0.00050	0.01520
Gender difference in labor force participation	0.00001	0.00810
Female primary school completion	0.13012	0.16920
Gender difference in primary school completion	0.08773	0.08820
Female secondary school completion	0.90692	0.92410
Gender difference in secondary school completion	0.71638	0.73140
Gender attitudes index	0.00225	0.05030
Gender attitudes index among women	0.01223	0.09620
Gender attitudes index among men	0.00063	0.03040

P-values estimated using 10,000 permutations, following procedures outlined in Section 5.5.2. For each outcome, the naïve p-value comes from the associated regression in a previous table. The permutation-based p-value is the fraction of permutations in which the magnitude of the estimated coefficient (from a hypothetical permutation of the gender indicator that respects the cluster structure of the language tree) exceeds the magnitude of the estimated coefficient in the true (non-permuted) data set. Distributions underlying first two rows are shown in Figure 6.

Table 8: Coefficient Stability

	OLS COEFFICIENTS			
	$\hat{\beta}$	$\tilde{\beta}$	$\beta^*(R_{max}, 1)$	δ^*
<i>Panel A. Cross-Country Regressions</i>				
Female labor force participation	-13.83	-11.92	-8.35	1.44
Gender difference in labor force participation	-11.61	-14.66	-17.87	3.24
Female primary school completion	14.79	-6.71	-19.40	$\delta < 0$
Gender difference in primary school	1.21	-3.72	-6.27	$\delta < 0$
Female secondary school completion	14.52	0.43	-9.69	0.05
Gender difference in secondary school	0.48	-0.86	-1.77	$\delta < 0$
Gender attitude index	-0.03	-0.12	-0.20	$\delta < 0$
Gender attitudes among women	-0.02	-0.10	-0.18	$\delta < 0$
Gender attitudes among men	-0.04	-0.14	-0.23	$\delta < 0$
<i>Panel B. Individual-Level Regressions — Afrobarometer Data</i>				
In labor force (Table A5, women only)	-0.24	-0.18	-0.13	2.11
Female \times in labor force (Table A6)	-0.17	-0.11	-0.06	1.86
Completed primary school (Table A7, women only)	-0.31	-0.22	-0.15	2.18
Female \times completed primary school (Table A8)	-0.12	-0.11	-0.10	4.64
Completed secondary school (Table A7, women only)	-0.19	-0.16	-0.14	3.47
Female \times completed secondary school (Table A8)	-0.06	-0.06	-0.06	6.01
<i>Panel C. Individual-Level Regressions — India IHDS Data</i>				
In labor force (Table A9, women only)	-0.08	-0.07	-0.07	11.70
Female \times in labor force (Table A10)	-0.10	-0.08	-0.04	1.90
Completed primary school (Table A11, women only)	-0.14	-0.13	-0.12	12.14
Female \times completed primary school (Table A12)	-0.13	-0.12	-0.11	13.19
Completed secondary school (Table A11, women only)	-0.03	-0.02	-0.02	7.20
Female \times completed secondary school (Table A12)	-0.03	-0.03	-0.03	25.89

Parameters estimated following procedures outlined in Altonji, Elder, and Taber (2005) and Oster (2017). $\hat{\beta}$ is the coefficient of interest from a bivariate regression. $\tilde{\beta}$ is the coefficient from a regression that includes the full set of observable controls. $\beta^*(R_{max}, 1)$ is the implied causal impact of grammatical gender on each outcome assuming a proportional selection coefficient (δ) equal to 1 and a maximum R^2 equal to 1.3 times the R^2 from the regression with controls (Oster 2017). δ^* is the proportional selection coefficient required to explain the observed relationship under the null hypothesis of no causal effect of grammatical gender on outcomes of interest.