I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13126

Gendered Language

Pamela Jakiela

Owen Ozier

APRIL 2020

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13126

Gendered Language

Pamela Jakiela
Center for Global Development, BREAD and IZA

Owen Ozier
World Bank Development Research Group, BREAD and IZA

APRIL 2020

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Gendered Language*

Languages use different systems for classifying nouns. *Gender languages* assign nouns to distinct sex-based categories, masculine and feminine. We construct a new data set, documenting the presence or absence of grammatical gender in more than 4,000 languages which together account for more than 99% of the world's population. We find a robust negative cross-country relationship between prevalence of gender languages and women's labor force participation and educational attainment. We replicate these associations in four countries in Sub-Saharan Africa and in India, showing that educational attainment and female labor force participation are lower among those whose native languages use grammatical gender.

JEL Classification: J16, Z10, Z13

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

Corresponding author:
Pamela Jakiela
Center for Global Development
2055 L St NW
Washington, DC 20036
USA
E-mail: pjakiela@cgdev.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, Quamrul Ashraf, Sarah Baird, Michal Bauer, Lori Beaman, Douglas Bernheim, Hoyt Bleakley, Premila and Satish Chand, Sameer Chand, Keith Chen, Julie Chytilová, Michael Clemens, Austin Davis, Stefano DellaVigna, Giles Dickenson-Jones, Penny Eckert, Benjamin Enke, 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 Hla, 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, Stelios Michalopoulos, Ted Miguel, Hannes Mueller, Nathan Nunn, 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), Duke, the European University Institute, Fordham, Georgetown, Harvard/MIT, 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.

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). Though linguistic determinism remains controversial (cf. McWhorter 2014), there is mounting evidence that 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). Russian speakers are better able to visually distinguish shades of blue than English speakers because Russian makes an obligatory distinction in shades that English does not (Winawer, Witthoft, Frank, Wu, Wade and Boroditsky 2007). 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 et al. 2003). Pérez and Tavits (2019) 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 of a link between grammatical gender and women's equality. To do this, we assemble 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; by bringing together linguists' work across these data sources, we construct 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, allowing us to make progress on previously intractable inference and contextual problems, as we discuss below.[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 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,

[1] WALS has also been used to study origins of language structures, as in Galor, Ozak and Sarid (2018).
[2] This calculation is based on *Ethnologue* estimates of the total number of native speakers in the world.

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 grammatical gender and female labor force participation. Our preferred specification suggests that grammatical gender is associated with a 9 percentage point reduction in women's labor force participation and a 10 percentage point increase in the gender gap in labor force participation. We also find a negative cross-country relationship between grammatical gender and women's educational attainment. Though women's labor force participation and educational attainment both increased substantially in recent decades, the negative association with grammatical gender is quite persistent over time. These empirical results are robust to the inclusion of a wide range of controls. Using data from the World Values Survey (WVS), we also show that grammatical gender predicts support for traditional gender roles among both men and women.

These correlations raise the question of whether language is associated with persistent, observable cultural features that predict labor force participation and educational attainment. To address this, we match languages to ethnographic groups in the most comprehensive available data source, George Murdock's Ethnographic Atlas (Murdock 1967). We use simple machine learning techniques to identify pre-industrial ethnographic characteristics that predict use of a gender language. We identify three: use of the plough, raising horses or camels, and regularly milking domestic animals. We include these cultural practices as controls throughout our analysis. Thus, we identify the association between grammatical gender and women's participation in economic life after controlling – to the extent possible – for likely cultural and historical confounds.

Our new data allow us to address two statistical concerns with cross-country analysis that previously available data could not. First, we observe our independent variable of interest at the country level, but only up to an interval, since there remains a small fraction of the population for whom we are uncertain of the status of the language they speak. Using a bounding technique proposed by Imbens and Manski (2004), we show that our thorough coverage of the world's languages produces estimates that are nearly unchanged when correcting for the interval nature of our independent variable of interest; this would not have been true with pre-existing datasets. Second, languages are not independent: within a language family, individual tongues have evolved in parallel over many centuries (Roberts, Winters and Chen 2015). While this slow process of language development may help to address potential concerns about reverse causality, it complicates statistical inference. Intuitively, linguistic characteristics vary across clusters of related languages, but individual countries draw from many different language clusters, making conventional clustering (of standard errors) impossible. We address this issue by implementing a permutation test, made possible by our novel data set, 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 below which we observe no 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 cross-country associations that we observe are not spurious.

To further examine the empirical link between gender languages and women's involvement in economic life, we examine the individual-level association between 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. Our new linguistic dataset makes this analysis possible: we characterize the grammatical gender structure of 352 languages spoken in Nigeria and 352 spoken in India, whereas the WALS includes only eight Nigerian and 23 Indian languages. 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 cultural 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.

To summarize, we document associations between grammatical gender and outcomes for women. We show that these associations are robust to controlling for pre-industrial cultural traits as well as other geographic and historical factors. It is clear that 21st-century labor market outcomes cannot have caused pre-industrial language characteristics, so reverse causality cannot explain the observed empirical relationship. This leaves two possibilities. One is that grammatical gender has a causal impact on 21st-century human behavior. The other is that both 21st-century human behavior and pre-industrial grammatical gender were caused by heretofore undocumented cultural features (not included in Murdock's *Ethnographic Atlas*). We cannot rule this out. However, such an alternative explanation runs counter to the widely held view that the structure of language is not empirically linked to culture in any meaningful way (McWhorter 2014).[3] Whether the underlying cause is language structure itself or some other unobserved pre-industrial cultural trait, it must explain why an empirical link is present within countries on two continents and across countries globally.

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 four thousand languages. Section 5 presents our cross-country anal-

[3] For example, in a recent book intended for a popular audience, the linguist John McWhorter writes: "The variety among the world's languages in terms of how they work is unrelated to the variety among the world's peoples, and thus Whorfianism cannot be saved even by fashioning a dynamic two-way relationship between cultures and the languages that they are spoken in" (McWhorter, 2014, p. 37).

ysis, 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).[4] 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.[5]

[4]There 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 agrement" 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.

[5]Corbett (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

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 1991, Aikhenvald 2003, Hellinger 2003).[6] 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*.[7] 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 "la casa blanca" because "house" is feminine, but "the white horse" is "el caballo blanco" 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.[8] Gender languages differ

---

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.

[6] Almost 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.

[7] Swahili, 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.

[8] Moreover, 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 et al. 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

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, 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).[9,10]

and females. Givati and Troiano (2012) show that countries where the dominant language makes pronominal gender distinctions have shorter government-mandated maternity leaves.

[9]In 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).

[10]No 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-

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 classifacatory fact in our thought-world" (Whorf 2011[1956]b, p. 69).[11] This argument — which Whorf advanced without offering any empirical evidence to support it — has been controversial (cf. McWhorter 2014). However, recent work in psychology and political science shows that grammatical gender shapes our subconscious attitudes in subtle and surprising ways. For example, Boroditsky et al. (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 et al. 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 (2019) conduct a

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 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.

[11]His 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).

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 were randomly assigned.[12]

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 et al. (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.[13] 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 as a comprehensive data source. 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 vari-

[12]There 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.

[13]In related work, Gay et al. (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.

able (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, 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.[14] 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. In our model, grammatical gender does not cause individuals to discriminate against women; instead, it predisposes people toward thinking in terms of separate masculine and feminine domains or spheres — thereby constraining the actions of both men and women.

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

[14] We discuss these clusters further in Section 5.5.2.

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.[15]

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 (i.e. the payoff associated with the binary decision to attend school in our simple model) 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 continuous cumulative density 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 continuous and 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

$$P^*_{girls} = \frac{1 - F_\gamma(\gamma^*)}{2 - F_\beta(\beta^*) - F_\gamma(\gamma^*)} \tag{1}$$

of students enrolled in school. The model is symmetric: if $F_\gamma = F_\beta$ and $R_g(\cdot) = R_b(\cdot)$, then

[15] To 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 model to consider the possibility that these costs exist for everyone but might be larger for those speaking gender languages, for whom partitioning the world into masculine, feminine, and potentially neutral spheres might appear more natural.

$\gamma^* = \beta^*$ and the equilibrium fraction of girls among enrolled students, λ^*, is $\frac{1}{2}$.

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) 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 the threshold value, 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 could 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.[16] 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.

[16]Fryer 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.

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 female/woman/mother i is characterized by $\gamma_i \sim F_\gamma$ and the ability of male/man/father is characterized by $\beta_i \sim F_\beta$.[17] γ and β both have continuous support between 0 and some finite maxima, β^{max} and γ^{max}.

A household maximizes consumption:

$$C = w_{mom}L_{mom} + w_{dad}L_{dad} - w_{nanny}H_{nanny} \tag{2}$$

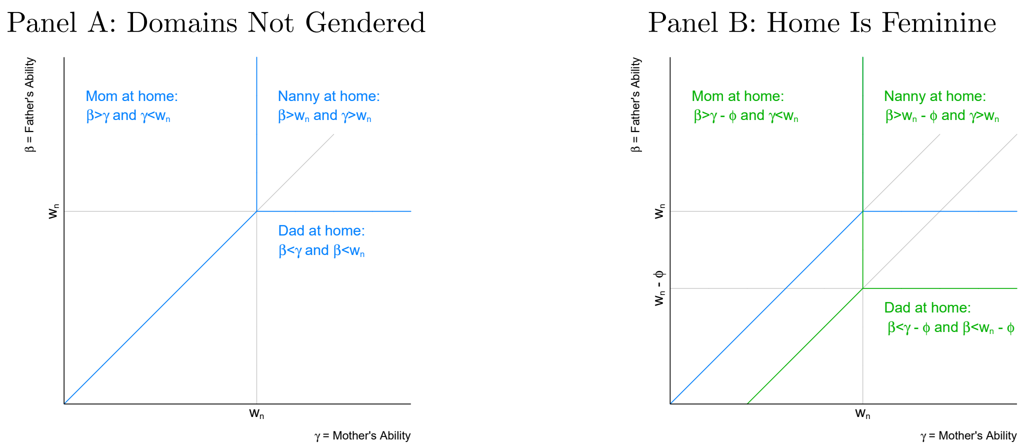
where $w_{mom} = \gamma_i$ indicates the wage that a mom of ability γ_i earns if she works outside the home, $w_{dad} = \beta_i$ the wage that a dad of ability β_i earns if he works outside the home, and w_{nanny} represents the market wage paid to nannies. We assume that nannies are female, and that 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).[18] 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.

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

[18] While 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 more realistic but potentially less helpful in illustrating the key results. As discussed in (Lancy 2015), childcare is either done by mothers or by other girls and women (including many older women) in most human societies.

Figure 1: Labor Force Participation in Two Equilibria



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:

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

In other words, P^*_{mom} is the integral of $f_{\beta,\gamma}(\beta,\gamma)$ over the "mom at home" region in Figure 1. P^*_{dad} and P^*_{nanny} can be defined analogously:

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

and

$$P^*_{nanny} = \int_{\beta=w_n}^{\beta=\beta^{max}} \int_{\gamma=w_n}^{\gamma=\gamma^{max}} f_{\beta,\gamma}(\beta,\gamma). \tag{5}$$

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:

$$\frac{P^*_{dad} + P^*_{nanny}}{1 + P^*_{nanny}} \tag{6}$$

since every household sends at least one adult into the workforce, and those with nannies send two. $\left(P^*_{mom}, P^*_{dad}, P^*_{nanny}\right)$ 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, 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, $P^{NN}_{dad} = P^*_{dad}$, $P^{NN}_{mom} = P^*_{mom}$, and $P^{NN}_{nanny} = P^*_{nanny}$. However, when domains can be gendered, this is only an equilibrium when

$$\lambda < P^{NN}_{mom} + P^{NN}_{nanny} < 1 - \lambda \tag{7}$$

and

$$\lambda < \frac{P^{NN}_{dad} + P^{NN}_{nanny}}{1 + P^{NN}_{nanny}} < 1 - \lambda. \tag{8}$$

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 as λ approaches one half, limiting the scope for non-gendered domains.

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

$$C = \gamma_i - \phi. \tag{9}$$

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.[19] Hence, many societies 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.

The model is symmetric, and it does not automatically predict that grammatical gender leads to worse outcomes for women and girls. Instead, it suggests that grammatical gender can exacerbate gender gaps that would result from small differences in, for example, the return to schooling. Importantly, our framework suggests that the tendency to partition the non-linguistic realm into masculine and feminine domains may influence the economic decisions of both men and women – for example, by limiting men's participation in the home or in other domains dominated by women in equilibrium.

[19] There 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 not imposed any structure.

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, 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.[20] To address this concern, we use two independent sources to characterize the grammatical gender structure of each language whenever possible. 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. All existing databases compiled by a single team of linguists using explicit, uniform standards to classify gender and non-gender languages cover far fewer languages.

4.1 Building a Grammatical Gender Data Set

Data on the set of all mother tongues 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 mother tongue. Of the 7,457 languages included in the *Ethnologue* database, we drop languages that are extinct or have no native speakers, sign languages, 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

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

6.44 billion native speakers (or more than 99 percent of the total population, according to the *Ethnologue*).

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 that characterize each language's grammatical gender structure) 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 (i.e. all nouns are assigned to classes that determine obligatory agreement) 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.[21] Second, whenever possible, we also code an indicator for dichotomous gender languages (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

[21] As 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.

of the world's population; we identify two independent sources that confirm the grammatical gender structure of 2,561 languages. 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), Iran (93.7 percent), Ethiopia (92.6 percent), 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.

Though more than a third of the world's population speaks a gender native language, only 441 languages (10.2 percent) use grammatical gender. This suggests that societies and cultures that use gender languages may not be representative of the set of all cultures – in other words, grammatical gender may not be plausibly exogenous. To explore this possibility, we merge our database of languages to the Ethnographic Atlas, anthropologist George Murdock's compilation of ethnographic work on pre-industrial societies (Murdock 1967). The Ethnographic Atlas characterizes the cultural practices of early societies on a range of dimensions including kin structures, food production, and gender norms. We identify the cultural practices that predict use of a gender language using lasso. Three such traits emerge: use of the plough, riding horses or camels, and regular milking of domestic animals. Importantly, cultural characteristics related to gender do not predict use of a gender language – though the significance of the plough is consistent with existing work (Boserup 1970, Alesina et al. 2013). We include language-level controls for early cultural practices that predict use of a gender language throughout our analysis.

4.2 Data from the Ethnographic Atlas

4.3 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 2015, which is available for 178 countries. We also use data on primary and secondary school completion in 2010 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 et al. (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.[22] 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.

[22] Kenya, 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.

5 Cross-Country Analysis

5.1 Empirical Strategy

In our cross-country analysis, the independent variable of interest is 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:

$$Y_c = \alpha + \beta Gender_c + \delta_{continent} + \lambda X_c + \varepsilon_c \tag{10}$$

where Y_c is the dependent variable in country c, $Gender_c$ 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 of country-level controls for geographic and ethnographic characteristics, and ε_c is a conditionally mean-zero error term.[23] Standard errors are clustered at the language level (by the most widely spoken language within each country).

Our main outcomes of interest are women's labor force participation and educational attainment. However, we do not wish to conflate cross-country differences in women's outcomes with structural factors that impact labor force participation and educational attainment among both men and women. To rule out this possibility, we also report specifications where the outcome variable is the gender gap calculated as the linear difference between women's and men's outcomes (e.g. women's labor force participation minus men's labor force participation). We also examine gender attitudes using data from the World Values Survey (WVS). 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 predicts support for traditional gender roles among both men and women.

[23] As 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 controls for geography and pre-industrial cultural practices, since these are plausibly exogenous.

5.2 Labor Force Participation

We examine the country-level relationship between grammatical gender and female labor force participation in Table 1. Women's labor force participation varies substantially across countries, from 9 percent in the Yemen to 87 percent in Madagascar. Table 1 demonstrates that female labor force participation is lower in countries where a larger fraction of the population has a gender mother tongue. In the first two columns, the outcome variable is the average level of female labor force participation in country c. Column 1 includes no controls. 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 9.4 percentage points higher in the absence of gender languages (p-value 0.003). Column 2 includes continent fixed effects plus additional controls for country-level geographic and ethnographic characteristics. The coefficient of interest is again negative and statistically significant. The coefficient suggests that grammatical gender is associated with a 9.3 percentage point decline in women's labor force participation (p-value 0.007).

In Columns 3 and 4 of Table 1, we replicate our analysis using the gender gap 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. In a specification with no controls (Column 3), we find that grammatical gender is associated with an 11.0 percentage point increase in the gender gap in labor force participation (p-value < 0.001). When we include controls (Column 4), grammatical gender is associated with a 10.2 percentage point increase in the gender difference in labor force participation (p-value 0.001). 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.[24]

[24] In the Online Appendix, we report a range of robustness checks, all of which suggest that the relationship between grammatical gender and female labor force participation is not driven by outliers or specification choices. We obtain similar results when we use the ratio of female labor force participation to male labor force participation as the outcome variable (Online Appendix Table A1). 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 et al. 2016). Nevertheless, results are broadly similar when we control for log GDP per capita, population, major world religions, and an indicator for post-Communist regimes. In Online Appendix Table A3, we

In Figure 3, we show that the association between grammatical gender and female labor force participation has been remarkably stable over the last 25 years – though female labor force participation has increased substantially.[25] The association between the proportion of a country's population speaking a gender native language and female labor force participation is negative and statistically significant in every year for which data is available, as is the relationship between grammatical gender and the gender gap in labor force participation. Thus, recent increases in women's labor supply have done little to weaken the empirical link between grammatical gender and women's economic activity.

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 educational attainment are not nearly as large as gender gaps in labor force participation. Across the 142 countries in the Barro-Lee data set, the median gender gap in educational attainment is less than half a year of schooling, whereas the median gender gap in labor force participation is over 17 percentage points. These small gender gaps in years of schooling reflect the very high rates of educational attainment in many parts of the world, and particularly among industrialized nations. A growing number of countries offer free primary and secondary education, and many have compulsory schooling laws which tend to reduce gender gaps in attainment.

In Table 2, we examine the cross-country relationship between grammatical gender and educational attainment. As expected, the relationship is *positive* and significant when continent controls are not included (Column 1) — reflecting the fact that educational attain-

show that results are similar 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.

[25]Systematic data on female labor force participation is available from the World Bank for every year since 1990.

ment highest in Europe, where gender languages are dominant. Once continent fixed effects are included (Column 2), the estimated association is negative and marginally statistically significant (p-value 0.058). In Columns 3 and 4, we examine the relationship between grammatical gender and the gender gap in educational attainment. A negative and statistically significant relationship is evident once continent fixed effects and additional controls are included (Column 4). Coefficient estimates suggest that grammatical gender is associated with a 0.6 year increase in the gender gap in years of schooling (p-value 0.026).[26]

The Barro-Lee Educational Attainment Dataset provides estimates of adult educational attainment at five-year intervals from 1950 through 2010. In Figure 4, we show how the relationship between grammatical gender and women's educational attainment has evolved over the last 60 years. Though women's educational attainment has increased dramatically since 1950, the gender gap in years of schooling has remained relatively constant (Evans, Akmal and Jakiela 2020). Whether one considers the level of female educational attainment or the gender gap in schooling,the cross-country association with grammatical gender has grown more pronounced in recent years. One possible explanation is that educational attainment was quite low among both men and women in 1950; as opportunities for men and women have expanded, the constraints imposed by psychological and cultural constraints on women's equality may become more apparent.

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 5, 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

[26]We report a range of robustness checks in Online Appendix Tables A5 to A10. Results are similar if we use the rate of (or the gender gap in) primary school completion as the dependent variable (Online Appendix Table A5). We do not observe a statistically significant association between grammatical gender and women's likelihood of completing secondary school (Online Appendix Table A6). Results are qualitatively similar if we calculate the gender gap in educational attainment as a ratio rather than a linear difference (Online Appendix Table A7), include a range of "bad controls" including current GDP per capita (Online Appendix Table A8), or omit countries where the most widely spoken language is English, Spanish, or Arabic (Table A9).

gender languages predicts responses to six of the eight WVS questions.

In Table 3, 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 geographic and ethnographic characteristics, 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 55^{th} percentile) and Trinidad and Tobago (at the 80^{th} 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 Columns 3 through 6 of Table 3, we show that — as expected — there is a negative association between the country-level prevalence of grammatical gender and gender attitudes among both women (Columns 3 and 4) and men (Columns 5 and 6). Though the coefficient is slightly larger for men, we cannot reject equality across genders. Thus, the cross-country evidence suggests that grammatical gender predicts gender differences in behavior, 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. 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 et al. 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's 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 remains 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. 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:

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

where $\bar{C}$ satisfies

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

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 probability above 0.95 in either worst-case scenario.

In Table 4, 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 et al. (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

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

families.[28]

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 6 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

[28]Much of the observed variation in grammatical gender is across language families: the intra-class correlation is 0.69. Statistical approaches that ignore this variation often lack the statistical power to reject large treatment effects.

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 100,000 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 5, 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, permutation test p-values suggest that the observed associations are unlikely to have occurred by chance: six of the seven estimated coefficients remain statistically significant at at least the 90 percent level. 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.[29] 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

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

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 et al. 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 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 et al. 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.[30] 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 and representing 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:

$$Y_i = \alpha + \beta Gender_i + \gamma Z_i + \varepsilon_i \tag{13}$$

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, age^2, a set of religion dummies, and language-level controls for ethnographic characteristics associated

[30]The 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.

with the use of grammatical gender), and ε_i is a mean-zero error term. In our analysis of the Afrobarometer data, we also include 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:

$$Y_i = \alpha + \beta Gender_i + \zeta Female_i + \mu Gender \times Female_i + \gamma Z_i + \varepsilon_i \tag{14}$$

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, religion, and ethnography controls. Throughout this 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 A11 through A22). 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 in both contexts. We see a more muted association between grammatical gender and secondary school completion

(Panel C of Figure 7), though results still suggest a negative relationship in both the African and the Indian data. 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 educational attainment, 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. We then find that the negative associations between grammatical gender and women's educational attainment and labor force participation are broadly 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.

Because grammatical structures evolve over many centuries, modern gender norms could not explain the observed empirical relationship between grammatical gender and women's equality. Instead, any alternative causal mechanism must involve some pre-modern cultural characteristic that could have shaped both linguistic structure and gender norms.[31] In our main analysis, we have attempted to address this possibility in two ways. First, we used simple machine learning techniques to identify the pre-modern cultural traits that predicted use of a gender language, combining our linguistic database with the Ethnographic Atlas (Murdock 1967); these variables are included as controls throughout our analysis. Second, we also control for religious affiliation to the extent possible – as a standard control in our within-country specifications and as a robustness check in our cross-country analysis.

To further 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).[32] 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 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 report the proportional selection coefficient, δ^*. 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.

Coefficient stability results are presented in Online Appendix Table A23; all seven cross-country specifications and 11 of 12 within-country specifications are sufficiently robust to

[31] Even such an alternative explanation runs counter to existing work in linguistics that suggests that grammatical structures are not culturally determined. See McWhorter (2014) for discussion.

[32] An 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.

the inclusion of controls to suggest that the observed association is unlikely to be explained by unobservables. Thus, the coefficient stability approach supports the hypothesis that grammatical gender has a causal impact on women's labor force participation and women's educational attainment both across and within countries. 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 straightforward causal interpretation 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 and educational attainment. At the country level, an increase in the proportion of the population whose native language is a gender language is associated with larger gender differences in labor force participation and schooling attainment. Using data from the World Values Survey, we show that grammatical gender also predicts support for traditional gender roles among both women and men. 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 empirical link between grammatical gender and women's equality exists within countries. Speaking a gender native language is associated with lower labor force participation and primary school completion among women, both in absolute terms and relative to men from the same ethnolinguistic group. Using data from the Ethnographic Atlas, we use machine learning to identify a set of pre-industrial cultural traits that predict the use of a gender languages. The three traits so identified – use of the plough, riding horses or camels, and regularly milking domestic animals – reflect the organization of food production rather than pre-modern gender dynamics. Moreover, both our cross-country and within-country regressions are robust to the inclusion of these cultural characteristics as controls – suggesting that the observed correlation between the use of grammatical gender and modern gender norms is not driven by any documented cultural characteristic.

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 may shape the way we think, and the way our children will think in the future.

References

Acharya, Avidit, Matthew Blackwell, and Maya Sen, "Explaining Causal Findings without Bias: Detecting and Assessing Direct Effects," *American Political Science Review*, 2016, *110* (3), 512–529.

Afrobarometer Data, "Kenya, Niger, Nigeria, Tanzania, Uganda, Rounds 2 through 5, 2002–2013," 2016. available at `http://www.afrobarometer.org`.

Aikhenvald, Alexandra Y., *Classifiers: A Typology of Noun Categorization Devices*, Oxford, UK: Oxford University Press, 2003.

Alesina, Alberto, Paola Giuliano, and Nathan Nunn, "On the Origins of Gender Roles: Women and the Plough," *Quarterly Journal of Economics*, 2013, *128* (2), 469–530.

Altonji, Joseph G, Todd E Elder, and Christopher R Taber, "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 2005, *113* (1), 151–184.

Angrist, Joshua D and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, 2008.

Aucejo, Esteban M, Federico A Bugni, and V Joseph Hotz, "Identification and inference on regressions with missing covariate data," *Econometric Theory*, 2017, *33* (1), 196–241.

Barro, Robert and Jong-Wha Lee, "A New Data Set of Educational Attainment in the World, 1950-2010," *Journal of Development Economics*, 2013, *104*, 184–198.

Beames, John, *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*, London, U.K.: Trubner and Company, 1875.

Boroditsky, Lera, Lauren A Schmidt, and Webb Phillips, "Sex, Syntax, and Semantics," in Susan Goldin-Meadow and Dedre Gentner, eds., *Language in Mind: Advances in the Study of Language and Thought*, Cambridge, MA: MIT Press, 2003, pp. 61–79.

Boserup, Esther, *Woman's Role in Economic Development*, London, UK: George Allen & Unwin, 1970.

Campbell, George L., *Compendium of the World's Languages*, London, UK: Routledge, 1991.

Chen, M. Keith, "The Effect of Language on Economic Behavior: Evidence from Savings Rates, Health Behaviors, and Retirement Assets," *Americn Economic Review*, 2013, *103* (2), 690–731.

Corbett, Greville G., *Gender*, Cambridge, UK: Cambridge University Press, 1991.

——, "Number of Genders," in Matthew S. Dryer and Martin Haspelmath, eds., *The World Atlas of Language Structures Online*, Leipzig, Germany: Max Planck Institute for Evolutionary Anthropology, 2013.

——, "Sex-based and Non-sex-based Gender Systems," in Matthew S. Dryer and Martin Haspelmath, eds., *The World Atlas of Language Structures Online*, Leipzig, Germany: Max Planck Institute for Evolutionary Anthropology, 2013.

——, "Systems of Gender Assignment," in Matthew S. Dryer and Martin Haspelmath, eds., *The World Atlas of Language Structures Online*, Leipzig, Germany: Max Planck Institute for Evolutionary Anthropology, 2013.

Creissels, Denis, "Typology," in Bernd Heine and Derek Nurse, eds., *African Languages: An Introduction*, Cambridge, UK: Cambridge University Press, 2000, pp. 231–258.

Danziger, Shai and Robert Ward, "Language Changes Implicit Associations between Ethnic Groups and Evaluation in Bilinguals," *Psychological Science*, 2010, *21* (6), 799–800.

Desai, Sonalde, Amaresh Dubey, and Reeve Vanneman, "India Human Development Survey-II (IHDS-II)," University of Maryland and National Council of Applied Economic Research, New Delhi 2015. distributed by: Inter-university Consortium for Political and Social Research, Ann Arbor, MI.

Durkheim, Emile and Marcel Mauss, *Primitive Classification*, Chicago, USA: University of Chicago Press, 1963. Translated from the French, edited, and with an introduction by Rodney Needham.

Evans, David K., Maryam Akmal, and Pamela Jakiela, "Gender Gaps in Education: The Long View," Center for Global Development Working Paper No. 523 2020.

Fryer, Roland G. and Steven D. Levitt, "An Empirical Analysis of the Gender Gap in Mathematics," *American Economic Journal: Applied Economics*, 2010, *2* (2), 210–40.

Galor, Oded, Omer Ozak, and Assaf Sarid, "Geographical Roots of the Coevolution of Cultural and Linguistic Traits," Working Paper 25289, National Bureau of Economic Research November 2018.

Gay, Victor, Daniel L. Hicks, Estefania Santacreu-Vasut, and Amir Shoham, "Decomposing culture: An Analysis of Gender, Language, and Labor Supply in the Household," *Review of Economics of the Household*, 2017, *16* (4), 879–909.

Givati, Yehonatan and Ugo Troiano, "Law, Economics, and Culture: Theory of Mandated Benefits and Evidence from Maternity Leave Policies," *Journal of Law and Economics*, 2012, *55* (2), 339–364.

Grant, Monica J. and Jere R. Behrman, "Gender Gaps in Educational Attainment in Less Developed Countries," *Population and Development Review*, 2010, *36* (1), 71–89.

Grierson, George Abraham, *Linguistic Survey of India: Volume V, Indo-Aryan Family Eastern Group, Part I, Specimens of the Bengali and Assamese Languages*, Calcutta, India: Superintendent Government Printing, 1903. `http://dsal.uchicago.edu/books/lsi/index.html` [accessed 13 July 2016].

——, *Linguistic Survey of India: Volume V, Indo-Aryan Family Eastern Group, Part II, Specimens of the Bihari and Oriya Languages*, Calcutta, India: Superintendent Government Printing, 1903. `http://dsal.uchicago.edu/books/lsi/index.html` [accessed 13 July 2016].

——, *Linguistic Survey of India: Volume VI, Indo-Aryan Family Mediate Group, Specimins of the Eastern Hindi Language*, Calcutta, India: Superintendent Government Printing, 1904. `http://dsal.uchicago.edu/books/lsi/index.html` [accessed 13 July 2016].

——, *Linguistic Survey of India: Volume VII, Indo-Aryan Family Southern Group, Specimins of theMarathi Language*, Calcutta, India: Superintendent Government Printing, 1905. `http://dsal.uchicago.edu/books/lsi/index.html` [accessed 13 July 2016].

——, *Linguistic Survey of India: Volume IX, Indo-Aryan Family Central Group, Part III, The Bhil Languages*, Calcutta, India: Superintendent Government Printing, 1907. `http://dsal.uchicago.edu/books/lsi/index.html` [accessed 13 July 2016].

——, *Linguistic Survey of India: Volume IX, Indo-Aryan Family Central Group, Part II, Specimens of Rajastani and Gujarati*, Calcutta, India: Superintendent Government Printing, 1908. `http://dsal.uchicago.edu/books/lsi/index.html` [accessed 13 July 2016].

——, *Linguistic Survey of India: Volume III, Tibeto-Burman Family, Part I, General Introduction, Specimins of the Tibetan Dialects, the Himalayan Dialects, and the North Assam Group*, Calcutta, India: Superintendent Government Printing, 1909. `http://dsal.uchicago.edu/books/lsi/index.html` [accessed 13 July 2016].

——, *Linguistic Survey of India: Volume IX, Indo-Aryan Family Central Group, Part I, Specimens of Western Hindi and Panjabi*, Calcutta, India: Superintendent Government Printing, 1916. `http://dsal.uchicago.edu/books/lsi/index.html` [accessed 13 July 2016].

——, *Linguistic Survey of India: Volume VIII, Indo-Aryan Family Northwestern Group, Part II, Specimens of the Dardic or Pisacha Languages*, Calcutta, India: Superintendent Government Printing, 1919. `http://dsal.uchicago.edu/books/lsi/index.html` [accessed 13 July 2016].

——, *Linguistic Survey of India: Volume X, Specimins of Languages of the Eranian Family*, Calcutta, India: Superintendent Government Printing, 1921. `http://dsal.uchicago.edu/books/lsi/index.html` [accessed 13 July 2016].

Guiora, Alexander Z, "Language and Concept Formation: A Cross-Lingual Analysis," *Cross-Cultural Research*, 1983, *18* (3), 228–256.

Hellinger, Marlis, "English," in Marlis Hellinger and Hadumod Bußman, eds., *Gender Across Languages: The Linguistic Representation of Women and Men, Volume*, Amsterdam, The Netherlands: John Benjamins Publishing Company, 2003, pp. 105–114.

—— **and Hadumod Bußman**, "The Linguistic Representation of Women and Men," in Marlis Hellinger and Hadumod Bußman, eds., *Gender Across Languages: The Linguistic Representation of Women and Men, Volume 3*, Amsterdam, The Netherlands: John Benjamins Publishing Company, 2003, pp. 1–26.

Hicks, Daniel L., Estefania Santacreu-Vasut, and Amir Shoham, "Does Mother Tongue Make for Women's Work? Linguistics, Household Labor, and Gender Identity," *Journal of Economic Behavior and Organization*, 2015, *110* (2), 19–44.

Horowitz, Joel L. and Charles F. Manski, "Censoring of outcomes and regressors due to survey nonresponse: Identification and estimation using weights and imputations," *Journal of Econometrics*, 1998, *84* (1), 37–58.

Imbens, Guido W. and Charles F. Manski, "Confidence Intervals for Partially Identified Parameters," *Journal of Econometrics*, 2004, *72* (6), 1845–1857.

Janhunen, Juha, "Grammaticl Gender from East to West," in Barbara Unterbeck and Matti Rissanen, eds., *Gender in Grammar and Cogniition*, Berlin, Germany: Mouton de Gruyter, 1999, pp. 689–708.

Kastovsky, Dieter, "Inflectional Classes, Morphological Restructuring, and the Dissolution of Old English Grammatical Gender," in Barbara Unterbeck and Matti Rissanen, eds., *Gender in Grammar and Cogniition*, Berlin, Germany: Mouton de Gruyter, 1999, pp. 709–728.

Kilarski, Marcin, *Nominal Classification: A History of Its Study from the Classical Period to the Present*, Amsterdam, The Netherlands: John Benjamins Publishing Company, 2013.

Krishnamurti, Bhadriraju, *Comparative Dravidian Linguistics*, Oxford, UK: Oxford University Press, 2001.

Lakoff, George, *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*, Chicago, IL: University of Chicago Press, 1987.

Lancy, David F., *The Anthropology of Childhood: Cherubs, Chattle, Changelings*, Cambridge, UK: Cambridge University Press, 2015.

Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig, eds.,, "Ethnologue: Languages of the World, nineteenth edition.," Dallas, Texas: SIL International 2016. Online version: `http://www.ethnologue.com` [accessed 4 June 2016].

Masica, Colin P., *The Indo-Aryan Languages*, Cambridge, UK: Cambridge University Press, 1991.

Mavisakalyan, Astghik, "Gender in Language and Gender in Employment," *Oxford Development Studies*, 2015, *43* (4), 403–424.

McWhorter, John H., *Defining Creole*, Oxford University Press, 2005.

____ , *The Language Hoax*, Oxford University Press, 2014.

Muhleisen, Susanne and Don E. Walicek, "Language and Gender in the Caribbean: An Overview," *Sargasso: Explorations of Language, Gender, and Sexuality*, 2010, *2008–2009* (I), 15–30.

Murdock, George Peter, "Ethnographic Atlas: A Summary," *Ethnology*, 1967, *6* (2), 109–236.

Ogunnaike, Oludamini, Yarrow Dunham, and Mahzarin R Banaji, "The Language of Implicit Preferences," *Journal of Experimental Social Psychology*, 2010, *46* (6), 999–1003.

Oster, Emily, "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business and Economic Statistics*, 2017.

Pérez, Efrén O. and Margit Tavits, "Language Influences Public Attitudes Toward Gender Equality," *Journal of Politics*, 2019, *81* (1), 81–93.

Roberts, Seán G., James Winters, and Keith Chen, "Future Tense and Economic Decisions: Controlling for Cultural Evolution," *PLOS One*, 2015, *10* (7), e0132145.

Santacreu-Vasut, Estefania, Amir Shoham, and Victor Gay, "Do female/male distinctions in language matter? Evidence from gender political quotas," *Applied Economics Letters*, 2013, *20* (5), 495–498.

____ **, Oded Shenkar, and Amir Shoham**, "Linguistic Gender Marking and Its International Business Ramifications," *Journal of International Business Studies*, 2014, *45*, 1170–1178.

Shoham, Amir and Sang Mook Lee, "The Causal Impact of Grammatical Gender Marking on Gender Wage Inequality amd Country Income Inequality," *Business and Society*, 2017.

Whorf, Benjamin Lee, "Language, Mind, and Reality," in John B. Caroll, ed., *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, Mansfield Centre, CT: Martino Publishing, 2011[1956], pp. 246–270.

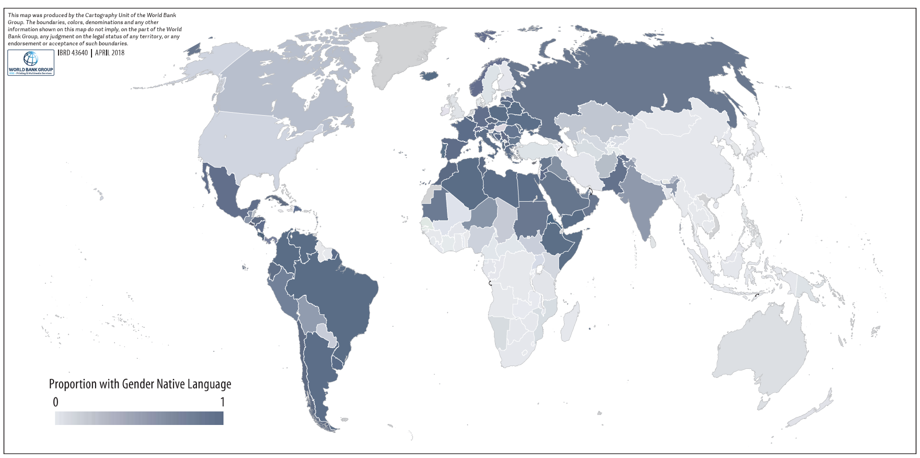
____ , "A Linguistic Consideration of Thinking in Primative Communities," in John B. Caroll, ed., *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, Mansfield Centre, CT: Martino Publishing, 2011[1956], pp. 65–86.

Winawer, Jonathan, Nathan Witthoft, Michael C Frank, Lisa Wu, Alex R Wade, and Lera Boroditsky, "Russian blues reveal effects of language on color discrimination," *Proceedings of the National Academy of Sciences*, 2007, *104* (19), 7780–7785.

World Values Survey Association, "World Values Survey Wave 6 2010–2014 Official Aggregate v.20150418," 2015. available online at `www.worldvaluessurvey.org`, aggregate file producer: Asep/JDS, Madrid SPAIN.

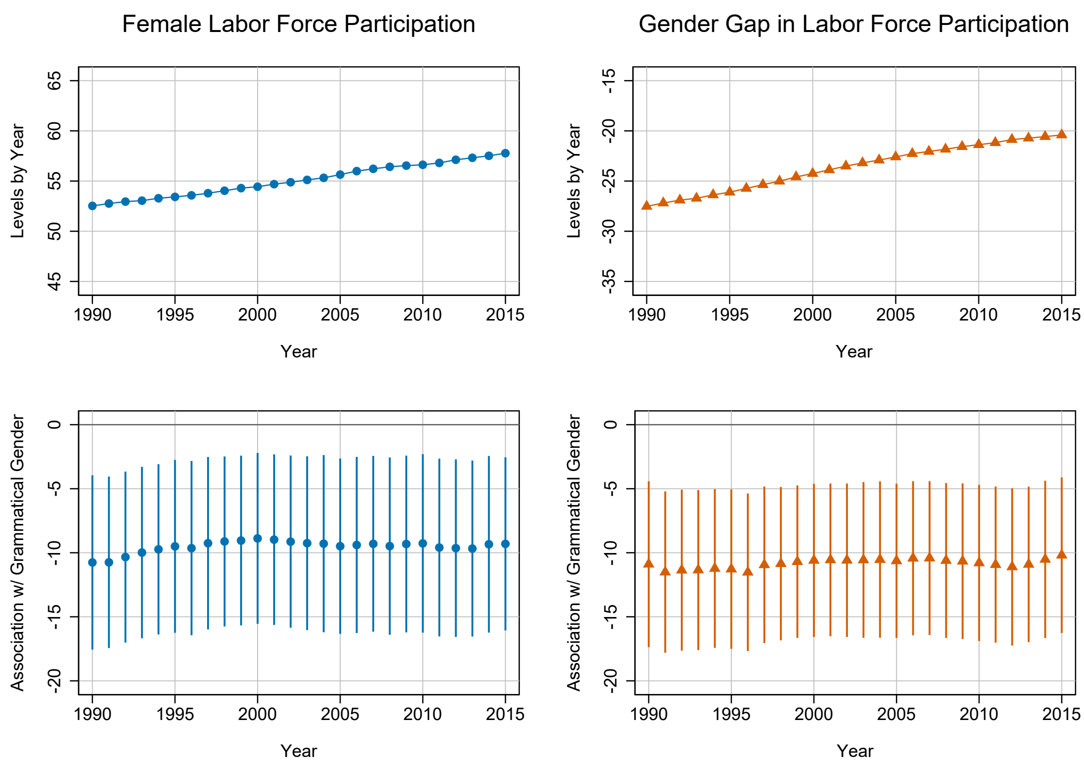
Young, Alwyn, "Consistency without Inference: Instrumental Variables in Practical Application," working paper 2018.

Figure 2: The Distribution of Gender Languages



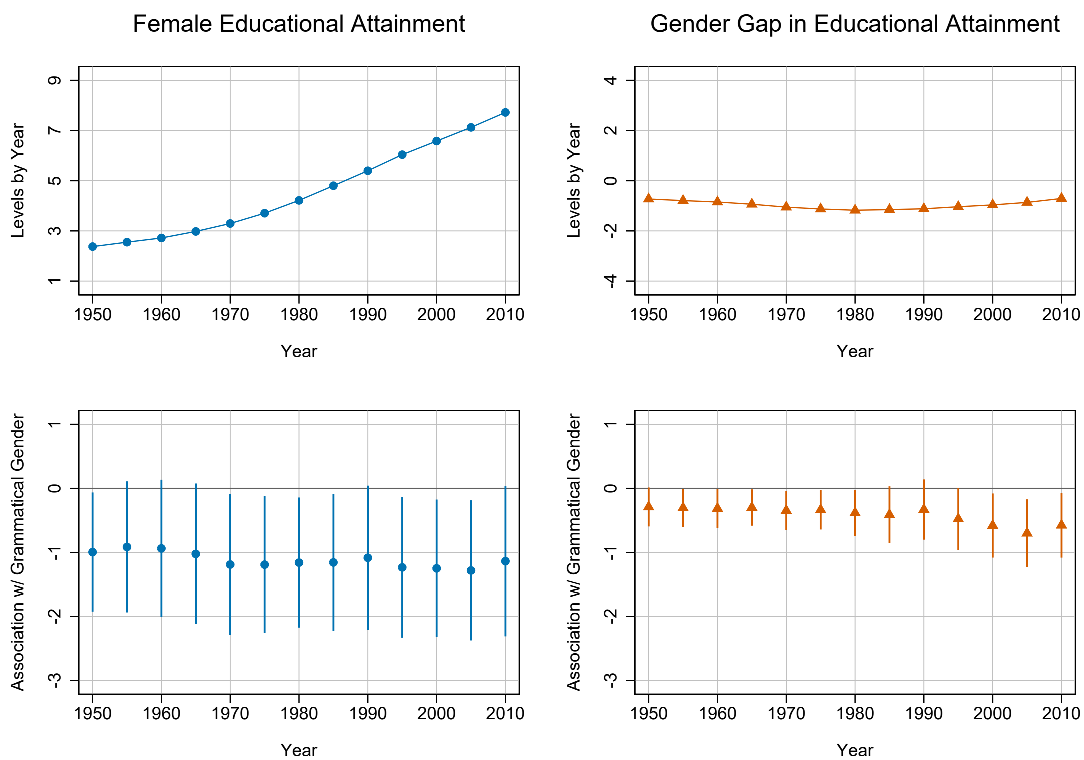
The figure shows the percentage of the native speakers in each country whose native language is a gender language (i.e. the fraction of *Ethnologue* native speakers whose native language uses a system of grammatical gender). The figure assumes that missing data (on 0.8 percent of all native speakers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

Figure 3: Labor Force Participation and Grammatical Gender Over Time



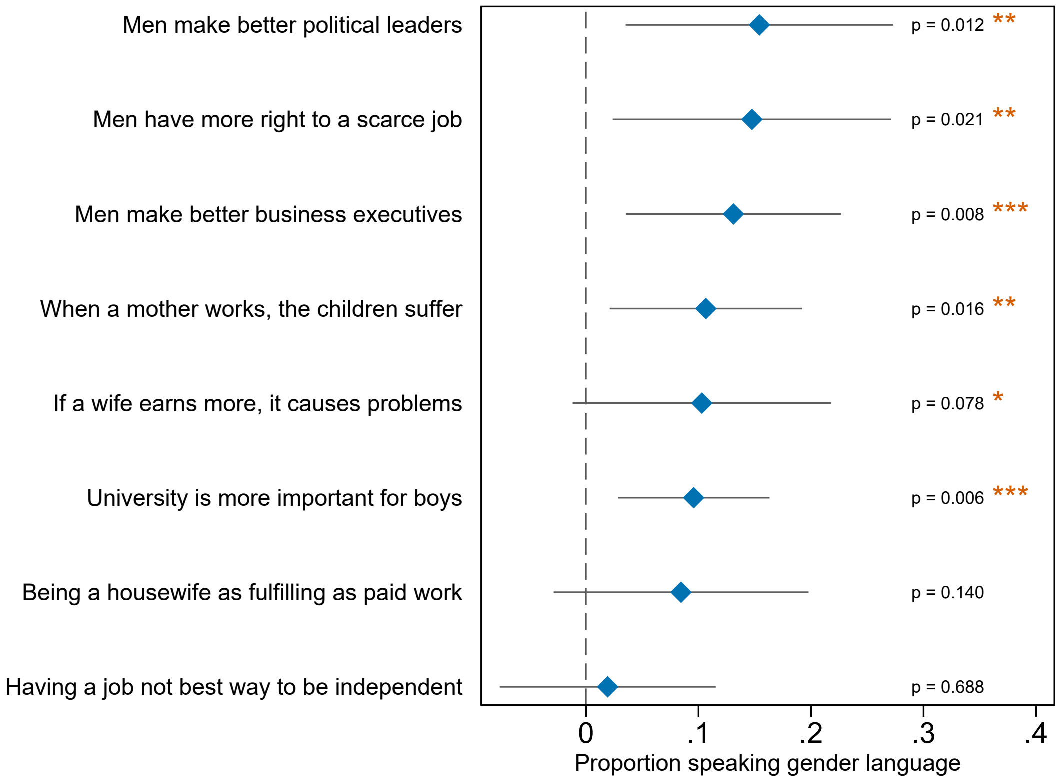
Data on labor force participation is available from the World Bank's World Development Indicators database for years 1990 through 2017. The gender gap in labor force participation is measured as the difference between female and male labor force participation. Lower panels report OLS coefficients and 95 percent confidence intervals from a regressions of labor force participation outcomes on our cross-country measure of the prevalence of gender native languages. Confidence intervals based on robust standard errors clustered by the most widely spoken language (by country). OLS specifications include continent fixed effects plus controls for the percentage of land area in the tropics or subtropics, average precipitation, average temperature, an indicator for being landlocked, and the Alesina *et al.* (2013) measure of suitability for the plough, and characteristics of pre-industrial societies (recorded at the ethnic group level in the Ethnographic Atlas) identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Figure 4: Educational Attainment and Grammatical Gender Over Time



Data on educational attainment among adults aged 15 and over is available from the Barro-Lee Educational Attainment Data Set at five-year intervals from 1950 through 2010. The gender gap in educational attainment is measured as the difference between female and male average years of schooling. Lower panels report OLS coefficients and 95 percent confidence intervals from a regressions of labor force participation outcomes on our cross-country measure of the prevalence of gender native languages. Confidence intervals based on robust standard errors clustered by the most widely spoken language (by country). OLS specifications include continent fixed effects plus controls for the percentage of land area in the tropics or subtropics, average precipitation, average temperature, an indicator for being landlocked, and the Alesina *et al.* (2013) measure of suitability for the plough, and characteristics of pre-industrial societies (recorded at the ethnic group level in the Ethnographic Atlas) identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Figure 5: 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 plus geography and ethnography controls. 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 6: Assignment to Clusters for the Permutation Test

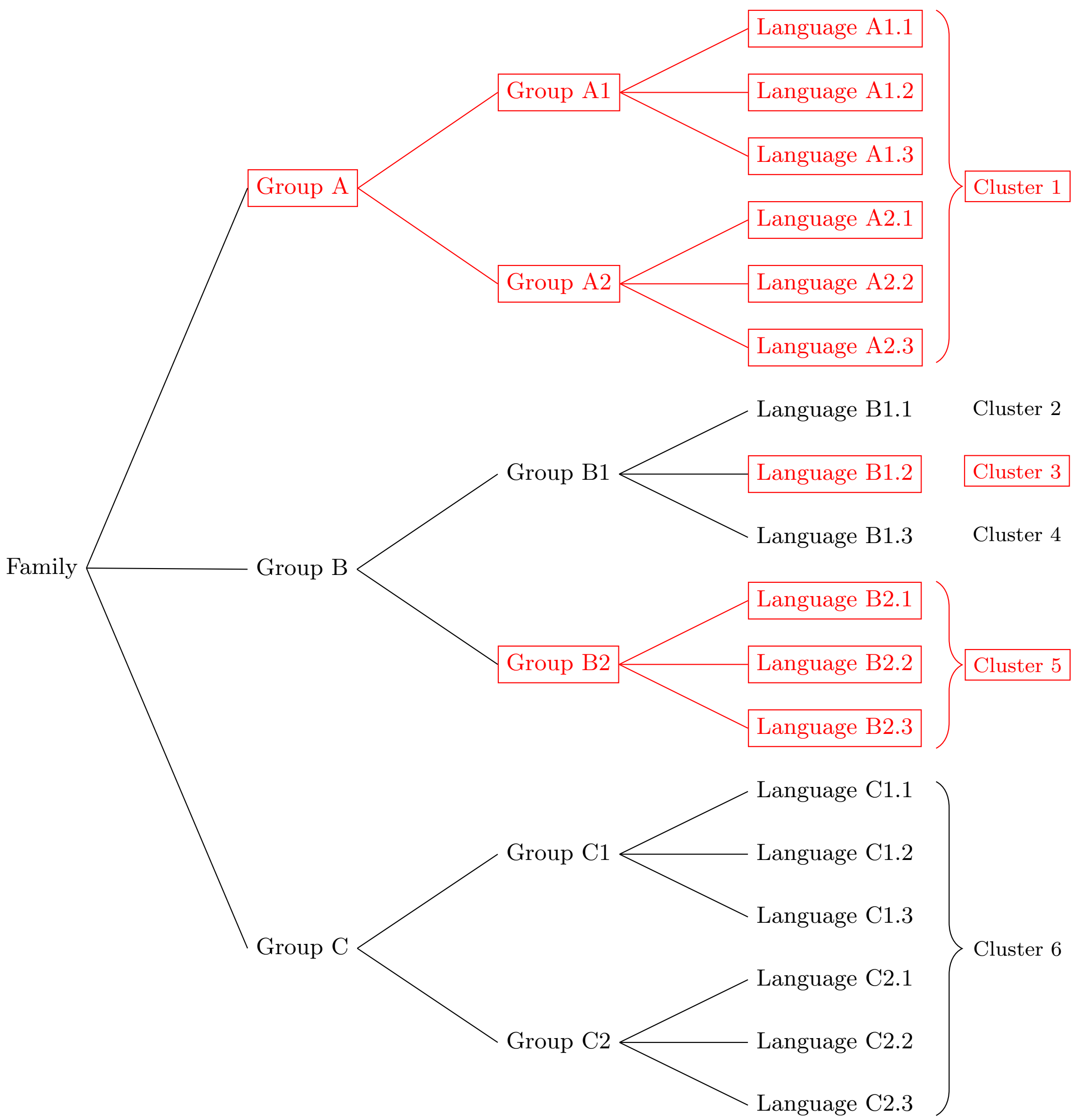
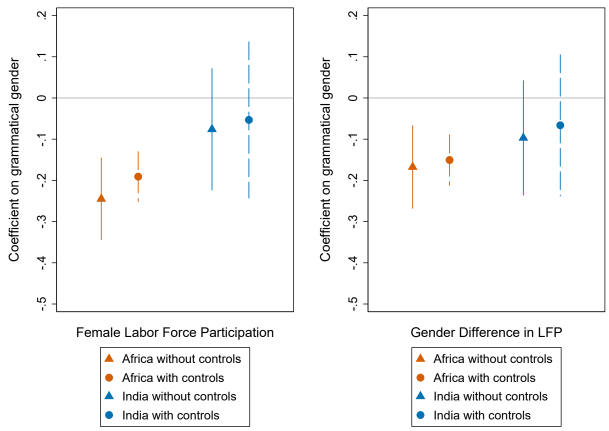


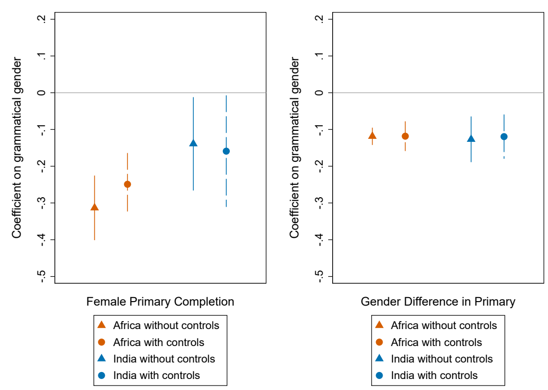
Figure illustrates a hypothetical language family. Gender languages and branches of the tree that include only gender languages are boxed and printed in red. Languages are assigned to clusters at the highest level of the language tree that shows no variation in grammatical gender.

Figure 7: Within-Country Variation in Grammatical Gender

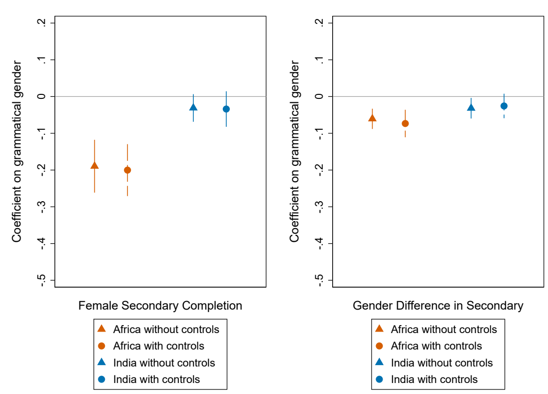
Panel A: Labor Force Participation



Panel B: Primary School Completion



Panel C: Secondary School Completion



Figures present OLS coefficients and confidence intervals from regressions of individual-level outcomes on an indicator for having a gender language as one's mother tongue. Data is from Rounds 2 through 5 of the Afrobarometer and from the India Human Development Survey-II. Controls are: age, age-squared, religion dummies, and ethnographic characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages.

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

Dependent variable:	LFP_f		LFP_f - LFP_m	
Specification:	OLS (1)	OLS (2)	OLS (3)	OLS (4)
Proportion gender	-9.40	-9.30	-11.02	-10.19
	(3.09)	(3.42)	(2.71)	(3.07)
	[0.003]	[0.007]	[$p < 0.001$]	[0.001]
Continent Fixed Effects	No	Yes	No	Yes
Geography/Ethnography Controls	No	Yes	No	Yes
Observations	178	178	178	178
R^2	0.06	0.37	0.09	0.53

Robust standard errors clustered by the most widely spoken language (by country) reported in parentheses. P-values are reported in square brackets. LFP_f is the percentage of women in the labor force, measured in 2015. LFP_f - LFP_m is the gender difference in labor force participation — i.e. the difference between female and male labor force participation. Country-level controls are the percentage of land area in the tropics or subtropics, average precipitation, average temperature, an indicator for being landlocked, and the Alesina *et al.* (2013) measure of suitability for the plough, and characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Table 2: Cross-Country OLS Regressions of Educational Attainment

Dependent variable:	EDU_f		EDU_f - EDU_m	
Specification:	OLS (1)	OLS (2)	OLS (3)	OLS (4)
Proportion gender	1.82	-1.14	0.25	-0.58
	(0.77)	(0.59)	(0.26)	(0.26)
	[0.019]	[0.058]	[0.346]	[0.026]
Continent Fixed Effects	No	Yes	No	Yes
Geography/Ethnography Controls	No	Yes	No	Yes
Observations	142	142	142	142
R^2	0.06	0.64	0.01	0.23

Robust standard errors clustered by the most widely spoken language (by country) reported in parentheses. P-values are reported in square brackets. EDU_f is educational attainment (years of schooling) among women aged 15 and over in 2010. EDU_f - EDU_m is the gender difference in educational attainment — i.e. the difference between female and male educational attainment, again measured in 2010. Country-level controls are the percentage of land area in the tropics or subtropics, average precipitation, average temperature, an indicator for being landlocked, and the Alesina *et al.* (2013) measure of suitability for the plough, and characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Table 3: Cross-Country OLS Regressions of Gender Attitudes

Dependent variable:	Gender Attitudes		Women's Attitudes		Men's Attitudes	
Specification:	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)	OLS (6)
Proportion gender	-0.03	-0.12	-0.02	-0.10	-0.04	-0.14
	(0.05)	(0.04)	(0.05)	(0.04)	(0.06)	(0.04)
	[0.576]	[0.005]	[0.714]	[0.020]	[0.508]	[0.002]

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 5) 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. Country-level controls are the percentage of land area in the tropics or subtropics, average precipitation, average temperature, an indicator for being landlocked, and the Alesina *et al.* (2013) measure of suitability for the plough, and characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Table 4: Robust Inference: Manski-Imbens Worst-Case 95-Percent Confidence Intervals

	Naïve OLS CI	Imbens-Manski CI
Female labor force participation	[−16.07, −2.54]	[−16.11, −2.43]
Gender difference in labor force participation	[−16.26, −4.11]	[−16.13, −3.61]
Female educational attainment	[−2.31, 0.04]	[−2.47, −0.02]
Gender difference in educational attainment	[−1.08, −0.07]	[−1.10, −0.09]
Gender attitudes index	[−0.20, −0.04]	[−0.20, −0.04]
Gender attitudes index among women	[−0.18, −0.02]	[−0.18, −0.02]
Gender attitudes index among men	[−0.22, −0.05]	[−0.22, −0.06]

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 5: Robust inference: Language structure

	Naïve OLS p-values	Permutation-based p-values
Female labor force participation	0.007	0.060
Gender difference in labor force participation	0.001	0.032
Female educational attainment	0.058	0.093
Gender difference in educational attainment	0.026	0.070
Gender attitudes index	0.005	0.076
Gender attitudes index among women	0.020	0.127
Gender attitudes index among men	0.002	0.045

P-values estimated using 100,000 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.

A Online Appendix: not for print publication

Table A1: Cross-Country Regressions of LFP Ratio

Dependent variable:	LFP_{ratio}	
Specification:	OLS (1)	OLS (2)
Proportion speaking gender language	-0.13	-0.12
	(0.04)	(0.04)
	[$p < 0.001$]	[0.002]
Continent Fixed Effects	No	Yes
Geography/Ethnography Controls	No	Yes
Observations	178	178
R^2	0.09	0.49

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_{ratio} is the ratio of the percentage of women in the labor force, measured in 2015, to the percentage of men in the labor force. Country-level controls are the percentage of land area in the tropics or subtropics, average precipitation, average temperature, an indicator for being landlocked, and the Alesina *et al.* (2013) measure of suitability for the plough, and characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Table A2: Cross-Country Regressions of LFP — Including “Bad” Controls

Dependent variable:	LFP_f	LFP_f - LFP_m
Specification:	OLS	OLS
	(1)	(2)
Proportion speaking gender language	-4.03	-6.78
	(3.15)	(2.77)
	[0.203]	[0.016]
Continent Fixed Effects	Yes	Yes
Geography/Ethnography Controls	Yes	Yes
Observations	173	173
R^2	0.63	0.71

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 2015. 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 2015. Country-level controls are the percentage of land area in the tropics or subtropics, average precipitation, average temperature, an indicator for being landlocked, and the Alesina *et al.* (2013) measure of suitability for the plough, and characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Table A3: Cross-Country Regressions of LFP — Dropping Major World Languages

Dependent variable:	LFP_f			$LFP_f - LFP_m$		
Omitted Language:	ARABIC	ENGLISH	SPANISH	ARABIC	ENGLISH	SPANISH
Specification:	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)	OLS (6)
Proportion speaking gender language	-3.80	-8.87	-8.00	-5.22	-10.24	-7.69
	(3.48)	(4.12)	(4.31)	(3.28)	(3.61)	(3.70)
	[0.277]	[0.033]	[0.066]	[0.114]	[0.005]	[0.040]
Continent Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Geography/Ethnography Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	159	167	160	159	167	160
R^2	0.21	0.39	0.41	0.37	0.55	0.56

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 2015. $LFP_f - LFP_m$ is the difference between male and female labor force participation in 2015. Country-level controls are the percentage of land area in the tropics or subtropics, average precipitation, average temperature, an indicator for being landlocked, and the Alesina *et al.* (2013) measure of suitability for the plough, and characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Table A4: Cross-Country Regressions of LFP — Weak vs. Strong Gender Categories

Dependent variable:	LFP_f		LFP_f - LFP_m	
Specification:	OLS (1)	OLS (2)	OLS (3)	OLS (4)
Proportion speaking (any) gender language	4.15	-6.23	6.38	-3.71
	(3.13)	(5.69)	(1.83)	(5.24)
	[0.187]	[0.276]	[$p < 0.001$]	[0.481]
Proportion speaking dichotomous gender language	-19.97	-4.63	-25.66	-9.76
	(5.53)	(6.45)	(4.08)	(5.84)
	[$p < 0.001$]	[0.474]	[$p < 0.001$]	[0.097]
Continent Fixed Effects	No	Yes	No	Yes
Geography/Ethnography Controls	No	Yes	No	Yes
Observations	178	178	178	178
R^2	0.16	0.37	0.28	0.54

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 2015. 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 2015. Country-level controls are the percentage of land area in the tropics or subtropics, average precipitation, average temperature, an indicator for being landlocked, and the Alesina *et al.* (2013) measure of suitability for the plough, and characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Table A5: Cross-Country OLS Regressions of Primary School Completion

Dependent variable:	PRI_f		PRI_f - PRI_m	
Specification:	OLS (1)	OLS (2)	OLS (3)	OLS (4)
Proportion gender	14.79	-9.40	1.21	-5.25
	(5.83)	(4.18)	(2.14)	(2.20)
	[0.013]	[0.027]	[0.573]	[0.019]
Continent Fixed Effects	No	Yes	No	Yes
Geography/Ethnography Controls	No	Yes	No	Yes
Observations	142	142	142	142
R^2	0.06	0.63	0.00	0.22

Robust standard errors clustered by the most widely spoken language (by country) reported in parentheses. P-values are reported in square brackets. PRI_f is the rate of primary school completion among women aged 15 and over in 2010. PRI_f - PRI_m is the gender difference in primary school completion, again measured in 2010. Country-level controls are the percentage of land area in the tropics or subtropics, average precipitation, average temperature, an indicator for being landlocked, and the Alesina *et al.* (2013) measure of suitability for the plough, and characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Table A6: Cross-Country OLS Regressions of Secondary School Completion

Dependent variable:	SEC_f		SEC_f - SEC_m	
Specification:	OLS (1)	OLS (2)	OLS (3)	OLS (4)
Proportion gender	14.52	-1.52	0.48	-3.05
	(5.77)	(3.92)	(1.93)	(2.02)
	[0.013]	[0.698]	[0.802]	[0.134]
Continent Fixed Effects	No	Yes	No	Yes
Geography/Ethnography Controls	No	Yes	No	Yes
Observations	142	142	142	142
R^2	0.06	0.68	0.00	0.17

Robust standard errors clustered by the most widely spoken language (by country) reported in parentheses. P-values are reported in square brackets. SEC_f is the rate of secondary school completion among women aged 15 and over in 2010. SEC_f - SEC_m is the gender difference in secondary school completion, again measured in 2010. Country-level controls are the percentage of land area in the tropics or subtropics, average precipitation, average temperature, an indicator for being landlocked, and the Alesina *et al.* (2013) measure of suitability for the plough, and characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Table A7: Cross-Country Regressions of Educational Attainment Ratio

Dependent variable:	EDU_{ratio}	
Specification:	OLS (1)	OLS (2)
Proportion speaking gender language	0.07	-0.10
	(0.04)	(0.04)
	[0.131]	[0.021]
Continent Fixed Effects	No	Yes
Geography/Ethnography Controls	No	Yes
Observations	142	142
R^2	0.02	0.30

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_{ratio} is the ratio of the percentage of women in the labor force, measured in 2015, to the percentage of men in the labor force. Country-level controls are the percentage of land area in the tropics or subtropics, average precipitation, average temperature, an indicator for being landlocked, and the Alesina *et al.* (2013) measure of suitability for the plough, and characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Table A8: Cross-Country Regressions of Education — Including “Bad” Controls

Dependent variable:	EDU_f	EDU_f - EDU_m
Specification:	OLS	OLS
	(1)	(2)
Proportion speaking gender language	-0.16	-0.30
	(0.57)	(0.28)
	[0.782]	[0.293]
Continent Fixed Effects	Yes	Yes
Geography/Ethnography Controls	Yes	Yes
Observations	140	140
R^2	0.75	0.42

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. EDU_f is educational attainment (years of schooling) among women aged 15 and over in 2010. EDU_f - EDU_m is the gender difference in educational attainment — i.e. the difference between female and male educational attainment, again measured in 2010. Country-level controls are the percentage of land area in the tropics or subtropics, average precipitation, average temperature, an indicator for being landlocked, and the Alesina *et al.* (2013) measure of suitability for the plough, and characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Table A9: Cross-Country Regressions of Education — Dropping Major World Languages

Dependent variable:	EDU_f			EDU_f – EDU_m		
Omitted Language:	Arabic	English	Spanish	Arabic	English	Spanish
Specification:	OLS	OLS	OLS	OLS	OLS	OLS
	(1)	(2)	(3)	(4)	(5)	(6)
Proportion speaking gender language	-0.96	-0.59	-0.86	-0.58	-0.48	-0.73
	(0.69)	(0.69)	(0.64)	(0.28)	(0.29)	(0.26)
	[0.165]	[0.396]	[0.182]	[0.042]	[0.101]	[0.006]
Continent Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Geography/Ethnography Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	127	132	124	127	132	124
R^2	0.66	0.62	0.66	0.29	0.19	0.24

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. EDU_f is educational attainment (years of schooling) among women aged 15 and over in 2010. EDU_f - EDU_m is the gender difference in educational attainment — i.e. the difference between female and male educational attainment, again measured in 2010. Country-level controls are the percentage of land area in the tropics or subtropics, average precipitation, average temperature, an indicator for being landlocked, and the Alesina *et al.* (2013) measure of suitability for the plough, and characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Table A10: Cross-Country Regressions of Educational Attainment — Weak vs. Strong Gender Categories

Dependent variable:	EDU_f		EDU_f - EDU_m	
Specification:	OLS	OLS	OLS	OLS
	(1)	(2)	(3)	(4)
Proportion speaking (any) gender language	4.98	0.34	0.39	-0.38
	(0.77)	(0.59)	(0.24)	(0.24)
	[$p < 0.001$]	[0.565]	[0.105]	[0.109]
Proportion speaking dichotomous gender language	-4.58	-2.39	-0.21	-0.32
	(0.61)	(0.65)	(0.32)	(0.26)
	[$p < 0.001$]	[$p < 0.001$]	[0.508]	[0.236]
Continent Fixed Effects	No	Yes	No	Yes
Geography/Ethnography Controls	No	Yes	No	Yes
Observations	142	142	142	142
R^2	0.20	0.66	0.01	0.24

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. EDU_f is educational attainment (years of schooling) among women aged 15 and over in 2010. EDU_f - EDU_m is the gender difference in educational attainment — i.e. the difference between female and male educational attainment, again measured in 2010. Country-level 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, and characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Table A11: OLS Regressions of African Women's Labor Force Participation

Dependent variable:	In Labor Force	
Specification:	OLS (1)	OLS (2)
Native language is a gender language	-0.24	-0.19
	(0.05)	(0.03)
	[$p < 0.001$]	[$p < 0.001$]
Country-Wave Fixed Effects	No	Yes
Individual Controls	No	Yes
Ethnography Controls	No	Yes
Observations	13154	13154
R^2	0.04	0.11

Robust standard errors clustered at the language level. The dependent variable is an indicator for being in the labor force (either working for a wage, self-employed, or actively seeking employment). Data is from Afrobarometer Rounds 2 through 5. The analysis includes data from Kenya, Niger, Nigeria, and Uganda; Niger was only added to the Afrobarometer in Round 5, while the other countries appear in all four rounds. Individual controls are age and age-squared and indicators for being identifying as Muslim, Catholic, Protestant, or another religion. Ethnographic controls are characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Table A12: OLS Regressions of Gender Differences in Labor Force Participation in Africa

Dependent variable:	In Labor Force	
Specification:	OLS (1)	OLS (2)
Female × gender language	-0.17	-0.15
	(0.05)	(0.03)
	[0.001]	[$p < 0.001$]
Native language is a gender language	-0.08	-0.04
	(0.02)	(0.02)
	[$p < 0.001$]	[0.011]
Female	-0.10	-0.15
	(0.01)	(0.03)
	[$p < 0.001$]	[$p < 0.001$]
Country-Wave Fixed Effects	No	Yes
Individual Controls	No	Yes
Ethnography Controls	No	Yes
Observations	26328	26328
R^2	0.04	0.12

Robust standard errors clustered at the language level. The dependent variable is an indicator for being in the labor force (either working for a wage, self-employed, or actively seeking employment). Data is from Afrobarometer Rounds 2 through 5. The analysis includes data from Kenya, Niger, Nigeria, and Uganda; Niger was only added to the Afrobarometer in Round 5, while the other countries appear in all four rounds. Individual controls are age and age-squared and indicators for being identifying as Muslim, Catholic, Protestant, or another religion, plus interactions between these controls and the female dummy. Ethnographic controls are characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Table A13: OLS Regressions of African Women's Primary School Completion

Dependent variable:	COMPLETE PRIMARY	
Specification:	OLS (1)	OLS (2)
Native language is a gender language	-0.31	-0.25
	(0.04)	(0.04)
	[$p < 0.001$]	[$p < 0.001$]
Country-Wave Fixed Effects	No	Yes
Individual Controls	No	Yes
Ethnography Controls	No	Yes
Observations	13142	13142
R^2	0.06	0.22

Robust standard errors clustered at the language level. The dependent variable is an indicator for completing primary school. Data is from Afrobarometer Rounds 2 through 5. The analysis includes data from Kenya, Niger, Nigeria, and Uganda; Niger was only added to the Afrobarometer in Round 5, while the other countries appear in all four rounds. Individual controls are age and age-squared and indicators for being identifying as Muslim, Catholic, Protestant, or another religion. Ethnographic controls are characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Table A14: OLS Regressions of Gender Differences in Primary School Completion in Africa

Dependent variable:	Complete Primary	
Specification:	OLS (1)	OLS (2)
Female × gender language	-0.12	-0.12
	(0.01)	(0.02)
	[$p < 0.001$]	[$p < 0.001$]
Native language is a gender language	-0.19	-0.12
	(0.04)	(0.03)
	[$p < 0.001$]	[$p < 0.001$]
Female	-0.08	-0.12
	(0.01)	(0.02)
	[$p < 0.001$]	[$p < 0.001$]
Country-Wave Fixed Effects	No	Yes
Individual Controls	No	Yes
Ethnography Controls	No	Yes
Observations	26294	26294
R^2	0.06	0.21

Robust standard errors clustered at the language level. The dependent variable is an indicator for completing primary school. Data is from Afrobarometer Rounds 2 through 5. The analysis includes data from Kenya, Niger, Nigeria, and Uganda; Niger was only added to the Afrobarometer in Round 5, while the other countries appear in all four rounds. Individual controls are age and age-squared and indicators for being identifying as Muslim, Catholic, Protestant, or another religion, plus interactions between these controls and the female dummy. Ethnographic controls are characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Table A15: OLS Regressions of African Women's Secondary School Completion

Dependent variable:	Complete Secondary	
Specification:	OLS (1)	OLS (2)
Native language is a gender language	-0.19	-0.20
	(0.04)	(0.04)
	[$p < 0.001$]	[$p < 0.001$]
Country-Wave Fixed Effects	No	Yes
Individual Controls	No	Yes
Ethnography Controls	No	Yes
Observations	13142	13142
R^2	0.02	0.15

Robust standard errors clustered at the language level. The dependent variable is an indicator for completing secondary school. Data is from Afrobarometer Rounds 2 through 5. The analysis includes data from Kenya, Niger, Nigeria, and Uganda; Niger was only added to the Afrobarometer in Round 5, while the other countries appear in all four rounds. Individual controls are age and age-squared and indicators for being identifying as Muslim, Catholic, Protestant, or another religion. Ethnographic controls are characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Table A16: OLS Regressions of Gender Differences in Secondary School Completion in Africa

Dependent variable:	Complete Secondary	
Specification:	OLS (1)	OLS (2)
Female × gender language	-0.06	-0.07
	(0.01)	(0.02)
	[$p < 0.001$]	[$p < 0.001$]
Native language is a gender language	-0.13	-0.13
	(0.04)	(0.03)
	[$p < 0.001$]	[$p < 0.001$]
Female	-0.08	-0.07
	(0.01)	(0.02)
	[$p < 0.001$]	[$p < 0.001$]
Country-Wave Fixed Effects	No	Yes
Individual Controls	No	Yes
Ethnography Controls	No	Yes
Observations	26294	26294
R^2	0.03	0.15

Robust standard errors clustered at the language level. The dependent variable is an indicator for completing secondary school. Data is from Afrobarometer Rounds 2 through 5. The analysis includes data from Kenya, Niger, Nigeria, and Uganda; Niger was only added to the Afrobarometer in Round 5, while the other countries appear in all four rounds. Individual controls are age and age-squared and indicators for being identifying as Muslim, Catholic, Protestant, or another religion, plus interactions between these controls and the female dummy. Ethnographic controls are characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Table A17: OLS Regressions of Indian Women's Labor Force Participation

Dependent variable:	In Labor Force	
Specification:	OLS (1)	OLS (2)
Native language is a gender language	-0.08	-0.05
	(0.07)	(0.10)
	[0.308]	[0.578]
Individual Controls	No	Yes
Ethnography Controls	No	Yes
Observations	39895	39895
R^2	0.01	0.04

Robust standard errors clustered at the language level. The dependent variable is an indicator for being in the labor force (reporting one's primary activity as agriculture, wage labor, self-employment, or salaried/professional work). Data is from India Human Development Survey-II (Desai, Dubey, and Vanneman 2015). Individual controls are age and age-squared and indicators for being identifying as Muslim, Christian, Sikh, or another religion. Ethnographic controls are characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Table A18: OLS Regressions of Gender Differences in Labor Force Participation in India

Dependent variable:	In Labor Force	
Specification:	OLS (1)	OLS (2)
Female × gender language	-0.10	-0.07
	(0.07)	(0.09)
	[0.171]	[0.444]
Native language is a gender language	0.02	0.01
	(0.01)	(0.01)
	[0.131]	[0.357]
Female	-0.56	-0.07
	(0.05)	(0.09)
	[$p < 0.001$]	[0.444]
Individual Controls	No	Yes
Ethnography Controls	No	Yes
Observations	75966	75966
R^2	0.40	0.47

Robust standard errors clustered at the language level. The dependent variable is an indicator for being in the labor force (reporting one's primary activity as agriculture, wage labor, self-employment, or salaried/professional work). Data is from India Human Development Survey-II (Desai, Dubey, and Vanneman 2015). Individual controls are age and age-squared and indicators for being identifying as Muslim, Christian, Sikh, or another religion. Ethnographic controls are characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Table A19: OLS Regressions of Indian Women's Primary School Completion

Dependent variable:	Complete Primary	
Specification:	OLS (1)	OLS (2)
Native language is a gender language	-0.14	-0.16
	(0.06)	(0.08)
	[0.033]	[0.040]
Individual Controls	No	Yes
Ethnography Controls	No	Yes
Observations	39895	39895
R^2	0.02	0.09

Robust standard errors clustered at the language level. The dependent variable is an indicator for completing primary school. Data is from India Human Development Survey-II (Desai, Dubey, and Vanneman 2015). Individual controls are age and age-squared and indicators for being identifying as Muslim, Christian, Sikh, or another religion. Ethnographic controls are characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Table A20: OLS Regressions of Gender Differences in Primary School Completion

Dependent variable:	Complete Primary	
Specification:	OLS (1)	OLS (2)
Female × gender language	-0.13	-0.12
	(0.03)	(0.03)
	[$p < 0.001$]	[$p < 0.001$]
Native language is a gender language	-0.01	-0.04
	(0.04)	(0.06)
	[0.767]	[0.496]
Female	-0.11	-0.12
	(0.02)	(0.03)
	[$p < 0.001$]	[$p < 0.001$]
Individual Controls	No	Yes
Ethnography Controls	No	Yes
Observations	75966	75966
R^2	0.05	0.10

Robust standard errors clustered at the language level. The dependent variable is an indicator for completing primary school. Data is from India Human Development Survey-II (Desai, Dubey, and Vanneman 2015). Individual controls are age and age-squared and indicators for being identifying as Muslim, Christian, Sikh, or another religion. Ethnographic controls are characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Table A21: OLS Regressions of Indian Women's Secondary School Completion

Dependent variable:	Complete Secondary	
Specification:	OLS (1)	OLS (2)
Native language is a gender language	-0.03	-0.03
	(0.02)	(0.02)
	[0.103]	[0.165]
Individual Controls	No	Yes
Ethnography Controls	No	Yes
Observations	39895	39895
R^2	0.00	0.02

Robust standard errors clustered at the language level. The dependent variable is an indicator for completing secondary school. Data is from India Human Development Survey-II (Desai, Dubey, and Vanneman 2015). Individual controls are age and age-squared and indicators for being identifying as Muslim, Christian, Sikh, or another religion. Ethnographic controls are characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Table A22: OLS Regressions of Gender Differences in Secondary School Completion

Dependent variable:	COMPLETE SECONDARY	
Specification:	OLS (1)	OLS (2)
Female × gender language	-0.03	-0.03
	(0.01)	(0.02)
	[0.027]	[0.129]
Native language is a gender language	0.00	-0.01
	(0.01)	(0.02)
	[0.957]	[0.662]
Female	-0.05	-0.03
	(0.01)	(0.02)
	[$p < 0.001$]	[0.129]
Individual Controls	No	Yes
Ethnography Controls	No	Yes
Observations	75966	75966
R^2	0.02	0.03

Robust standard errors clustered at the language level. The dependent variable is an indicator for completing secondary school. Data is from India Human Development Survey-II (Desai, Dubey, and Vanneman 2015). Individual controls are age and age-squared and indicators for being identifying as Muslim, Christian, Sikh, or another religion. Ethnographic controls are characteristics of pre-industrial societies identified by lasso as predictors of the use of gender languages (use of horses and/or camels, use of the plough, and regular milking of domestic animals).

Table A23: Coefficient Stability

	OLS Coefficients		
	$\mathring{\beta}$	$\tilde{\beta}$	δ^*
Panel A. Cross-Country Regressions			
Female labor force participation	-9.40	-9.30	2.05
Gender difference in labor force participation	-11.02	-10.19	1.84
Female educational attainment	1.82	-1.14	$\delta < 0$
Gender difference in education	0.25	-0.58	$\delta < 0$
Gender attitude index	-0.03	-0.12	$\delta < 0$
Gender attitudes among women	-0.02	-0.10	$\delta < 0$
Gender attitudes among men	-0.04	-0.14	$\delta < 0$
Panel B. Individual-Level Regressions — Afrobarometer Data			
In labor force (women-only specification)	-0.24	-0.19	1.42
In labor force (DD specification)	-0.17	-0.15	2.07
Completed primary school (women-only specification)	-0.31	-0.25	1.55
Completed primary school (DD specification)	-0.12	-0.12	2.89
Completed secondary school (women-only specification)	-0.19	-0.20	2.66
Completed secondary school (DD specification)	-0.06	-0.07	5.81
Panel C. Individual-Level Regressions —India Human Development Survey Data			
In labor force (women-only specification)	-0.08	-0.05	3.07
In labor force (DD specification)	-0.10	-0.07	0.55
Completed primary school (women-only specification)	-0.14	-0.16	12.82
Completed primary school (DD specification)	-0.13	-0.12	4.44
Completed secondary school (women-only specification)	-0.03	-0.03	11.29
Completed secondary school (DD specification)	-0.03	-0.03	2.45

Parameters estimated following procedures outlined in Altonji et al. (2005) and Oster (2017). $\mathring{\beta}$ 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. 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.