Gendered Language

Pamela Jakiela and Owen Ozier*

January 26, 2022

Abstract

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 relationship between prevalence of gender languages and women's labor force participation and educational attainment both across and within countries. We also demonstrate that grammatical gender is associated with both weaker legal support for women's equality and reduced female bargaining power within the household.

JEL codes: J16, Z10, Z13

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

*Jakiela: Williams College, BREAD, IZA, and J-PAL, email: pamela.jakiela@williams.edu; Ozier (corresponding author): Williams College, BREAD, IZA, and J-PAL email: oo3@williams.edu. 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, 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.

1 Introduction

Language structures our thoughts. 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).[1] Though linguistic determinism remains controversial (cf. McWhorter 2014, Fabb 2016), 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 et al. 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

[1] The theory of linguistic determinism – which posits that language structure determines thought patterns – is typically referred to as the Sapir-Whorf hypothesis. It was named after the anthropologist and linguist Edward Sapir and the linguist Benjamin Lee Whorf, though Whorf was its main proponent. Existing evidence does not support the strongest versions of linguistic determinism: language structure does not place hard limits on the human ability to think or innovate in any particular direction. Current research focuses on *linguistic relativism*, the idea that language structure can influence thought patterns at the subconscious level, potentially acting as a nudge or making certain heuristics more natural.

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* that assigns nouns (as opposed to people or other living beings with a biological sex) to masculine and feminine grammatical categories. 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 practice of partitioning the set of all nouns into distinct gender categories likely made other sex-based partitions 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, 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. 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 2018). 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 more than one third of the world's population is ignorable, yielding a set of bounding and clustering problems that severely hamper inference.[2] 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.[3]

Using this data set, we calculate — for every country in the world — an estimate of the proportion of the population whose native language is a gender language. 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 support for traditional gender roles. 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 associations with grammatical gender are quite persistent over time, and are robust to the inclusion of a wide range of controls.

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

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

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 country-level cultural characteristics 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. Linguistic characteristics vary between 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

100,000 counterfactual assignments of cluster-level grammatical gender 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.

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 Kenya, Niger, Nigeria, and Uganda and, separately, using the India Human Development Survey, which covers 33 Indian states. Our new linguistic dataset makes this analysis possible: we characterize the grammatical gender structure of 344 languages spoken in Nigeria and 352 spoken in India, whereas the WALS includes only 10 Nigerian and 29 Indian languages. Combining our language data with these surveys, 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 an individual's gender (i.e. the indicator for being female) and religious affiliation.

To summarize, we document associations between grammatical gender and adherence to traditional gender roles. 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.[4] This leaves two possibilities. One is that grammatical gender has a causal impact on 21st-century human be-

[4]Indeed, Tabellini (2008) argues that language can be used as an instrument for present-day cultural values because language structures evolved thousands of years ago, and any confounding impacts of ancestral culture on language occurred in the distant past, in many case among populations with only weak genetic and cultural ties to present-day speakers.

havior. 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).[5] 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.

A related question is *how* grammatical gender might impact women's equality. We outline a simple behavioral model illustrating how a predisposition toward viewing the world in terms of a dichotomy between masculine and feminine might strengthen the preference for adhering to traditional gender roles. However, since languages evolve over many many generations, any factor that increases the taste for sex-based household specialization could, over time, contribute to the evolution of informal and even formal institutions that limit women's outside options by discouraging women from involving themselves in economic life outside the home. To explore this mechanism, we examine the association between grammatical gender and two measures of women's equality that are not obviously linked to household specialization: laws protecting women's equality and women's decision-making power within the household. We observe a robust negative relationship netween grammatical gender and both of our measures of women's equality. Hence, grammatical gender is not simply a psychological nudge toward separate spheres; instead, it appears to be linked to present-day institutions that restrict women's autonomy and limit their legal rights.

The rest of this paper is organized as follows. Section 2 introduces the concept of grammatical gender, surveys recent research on its potential impacts, and discusses the channels through which grammatical gender might impact women's equality. Section 3 provides an overview of our data sources, including the data we have compiled on the grammatical structure of more than four thousand languages. Section 4 presents our cross-country and within-country analy-

[5]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).

sis of the relationship between grammatical gender and adherence to traditional gender norms. Section 5 explores the relationship between grammatical gender and measures of women's legal equality and decision-making power within the household. Section 6 concludes.

2 Conceptual Framework

2.1 Grammatical Gender

Many languages partition the set of all nouns into mutually exclusive categories that are referred to as either *genders* or *noun classes* (Corbett 1991, Aikhenvald 2003). Though nouns in the same class are often semantically related or morphologically similar, 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).[6] 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 nouns referring to inanimate objects are characterized as "masculine" while others are "feminine" (Aikhenvald 2003, Hellinger 2003). Following Aikhenvald (2003) and Hellinger and Bußman (2003), we say that a language uses a system of *grammatical gender* if (i) every noun belongs to a exactly one obligatory agreement class, (ii) the set of agreement classes includes masculine and feminine classes, and (iii) membership in the masculine and feminine agreement classes is a property of the noun itself, and not simply a reflection of the natural gender of an

[6] In Swahili, for example, the noun class determines the prefixes used to modify adjectives, verbs, demonstratives, and other parts of speech. For example, the word "chairs" (*viti*) belongs to the KI-/VI- class, while the word "teachers" (*walimu*) belongs to the M-/WA- class. Most adjectives used to modify "chairs" must therefore take the pre-fix *vi-*, but the same adjectives would typically begin with the prefix *wa-* when used to modify "teachers." So, "new chairs" is *viti vipya*, with the adjective "new" (*-pya*) taking the *vi-* prefix, but "new teachers" is *walimu wapya* because the adjective "new" instead takes the *wa-* prefix. Though most nouns in the KI-/VI- class begin with *ki-* in singular and *vi-* in plural (or *ch-* in singular and *vy-* in plural), it is the pattern of agreement with adjectives, verbs, etc. that defines the class. 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 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 a noun class system from other phonological or orthographic partitions of the set of all nouns.

animate referent.[7] We refer to languages characterized by such systems of grammatical gender as *gender languages*. Though all languages make human gender distinctions salient in some way (for example, by using "Mrs." rather than "Mr." to refer to a married person in polite conversation), grammatical gender systems extend the dichotomy between the human genders that is present in all languages to the realm of inanimate nouns, and to objects (e.g. a table or the moon) that do not have a (non-grammatical) gender or biological sex.[8] 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 the word "house" is feminine, but "the white horse" is "el caballo blanco" because the word "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 – for example, in the extent of agreement across parts of speech, and in 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

[7]Linguists disagree as to whether requiring "anaphoric agreement" between nouns that refer to humans and associated personal 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 (in which the grammatical gender that determines agreement is a property of the noun itself, and not simply a reflection of the gender or sex of the referent) 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. Employing the traditional definition of grammatical gender 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 referring to inanimate objects to gender categories) and systems that mark natural/human sex or gender morphologically.

[8]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). Because personal pronouns agree with the biological sex or 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 sex of animate referents (rather than the grammatical gender of the nouns themselves) are present in many languages that show no other form of gender inflection (Aikhenvald 2003, Creissels 2000). Other languages – e.g. Finnish, Hungarian, and Swahili – make no grammatical distinction between males and females (though different words are used for males and females in the same role – for example, a male sibling vs. a female sibling). Givati and Troiano (2012) show that countries where the dominant language makes pronominal gender distinctions have shorter government-mandated maternity leaves.

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).[9]

2.2 Related Literature

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).[10] 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) conduct 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

[9]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). 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]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).

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 a biological sex. Grammatical gender also appears to shape gender attitudes — even within individuals. Pérez and Tavits (2019) conduct a survey experiment with Estonian/Russian bilinguals, randomizing the language in which they are interviewed. They show that bilinguals who are interviewed in Russian (a gender language) are less supportive of gender equality than those who are interviewed in (non-gender) Estonian, even though interview languages were randomly assigned.[11]

Recent work also suggests that the influence of grammatical gender extends into the economic realm. Using the World Atlas of Language Structures (WALS), a catalog of 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. 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, while Gay et al. (2018) find that female immigrants to the United States exhibit lower labor market participation if they speak a gender language at home.

[11] 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.

2.3 Potential Causal Mechanisms

Existing work examining the relationship between grammatical gender and women's equality has not formally specified a potential causal mechanism. Whorf (2011[1956]b) suggests that the existence of a grammatical gender system that partitions nouns into masculine and feminine creates a subconscious predisposition to partition other categories in a similar manner – for example, dividing jobs into masculine and feminine occupations, or viewing public spaces as either women's or men's domains. In other words, the very existence of a grammatical gender structure might make sex a more salient dichotomy than other potential social distinctions such as age or class.[12]

How might a heightened predisposition to partition the world along gender lines manifest itself in society? We expect that it might lead to more rigid gender roles and greater segregation of tasks and domains by sex. A challenge inherent in testing such a theory is that the set of "feminine" or "masculine" tasks or domains differs across cultures, and has done so throughout human history. For example, weaving was a male task in some ancient cultures and a female task in others (Barber 1994). Agriculture was a male domain among the early Yoruba (Trigger 2003) and also among many cultures that used the plough (Boserup 1970, Alesina et al. 2013), but the Inka assigned men the task of digging holes and women the task of planting seeds in them (Trigger 2003).

As these examples illustrate, at least three issues must be considered when devising empirical tests of the association between grammatical gender and the extent to which tasks or domains are partitioned along gender lines. First, universality: one needs to identify domains or tasks that are relevant in all of the cultures that one wishes to study. For example, measures of the extent of gender segregation across occupational specializations (e.g. in obstetrics vs. surgery) or household tasks (e.g. who plows the fields or maintains the home wifi network) that make sense in one cultural context may not be appropriate in others. Second, when a sex-based

[12] Another possibility is that the gender of specific words (e.g. work, home) creates subconscious associations, along the lines suggested by (Boroditsky et al. 2003). While this mechanism is theoretically possible, most languages have many words referring to important domains like home or work; in gender languages, it is typically the case that some words referring to a domain or task are masculine while others are feminine (or neutral in languages with more than two genders). For example, to describe a firm in Spanish, one can either use the masculine word *negocio* or the feminine word *empresa*; to describe someone's participation in the labor market, one can use the masculine words *empleo* or *trabajo* or the feminine words *ocupación* or *carrera*.

partitioning of a domain exists, its direction (i.e. whether a role is considered predominantly masculine or feminine) must be common across cultures, or at least observable. As discussed above, weaving is an example of a task common in many early societies that was the exclusive purview of men in some cultures but women in others. Finally, even when groups or individuals face a common set of tasks, the level at which any sex-specific partitioning of roles occurs must be common, or (again) at least observable. As discussed above, while some societies treated agricultural as an entirely masculine domain, the Inka partitioned roles even within agriculture – but this does not necessarily mean that the strength of the norms restricting men and women to their sex-specific roles were any less strong among the Inka than elsewhere.

Household specialization – and in particular the implications of household specialization for women's participation in the labor market (Becker 1981, Becker 1985, Lundberg and Pollak 1993) – is a leading example of a setting where these three requirements are met, and one that has been studied by scholars interested in understanding the potential effects of grammatical gender (Hicks et al. 2015, Gay et al. 2018). Only women are capable of giving birth to babies, and women do most of the childcare in all human societies (?). Becker (1981) famously argues that women's biological comparative advantage in childbirth translates into an economic justification for "efficient" household specialization, with women specializing in domestic work and childrearing while men specialize in activities outside the home — particularly wage labor. Though Becker's model explains household specialization in the modern era, when at least one adult family member is typically engaged in paid work outside the home, sex-based specialization predates the industrial revolution (Barber 1994, Trigger 2003); Brown (1970) argues that women in preindustrial societies tended to be responsible for tasks that were compatible with childminding while men handled tasks that took them far from home and those that could not be done safely while children were present. Lundberg and Pollak (1993) characterize the equilibrium (which they term "separate spheres") in which women provide "household services" and men are responsible for earning an income in the labor market as "traditional gender roles" that are "generally recognized and accepted" (Lundberg and Pollak, 1993, pp. 993–994). Thus, both economists and other social scientists view the existence of separate spheres – with the home being the female sphere while the world outside the home, and particularly the labor market, is the male sphere – as a universally recognized example of a sex-based partition of the world into distinct male and female domains. However, the extent to which modern households adhere to these traditional gender roles varies substantially across societies.

In the Online Appendix, we provide a simple stylized model that illustrates how a more pronounced tendency to view domains as either masculine or feminine, which may result from speaking a gender language as one's mother tongue, could operate in the context of household specialization. The model incorporates a key insight of Akerlof and Kranton (2000): that entering domains that do not align with one's identity entails psychic and social utility costs. In our simplified model, a stronger tendency to view domains as either masculine or feminine increases the likelihood that either the home or the workplace is perceived as the domain of a single sex. The model is symmetric – there is no in-built assumption that women have a comparative advantage in domestic work – and the gender norms that can arise within the model constrain both women who could earn a high wage in the labor market and men who are relatively better suited to staying at home. Indeed, an important implication of the model is that whenever one of the two domains (home or work) is masculine, the other domain is feminine – so the same psychological predisposition that nudges women out of the labor market nudges men away from being stay-at-home parents. Nevertheless, the model predicts that when the tendency to partition the world into masculine and feminine domains is sufficiently strong, an equilibrium will exist in which work is a male domain while home is a female one.

Guided by the predictions of the model and building on prior empirical work, we hypothesize that grammatical gender may make the distinction between the separate spheres less malleable, reinforcing traditional gender norms and reducing the likelihood that women enter the labor force. We test this by examining both the cross-country and the within-country association between grammatical gender (specifically, whether one's native language uses a grammatical gender system) and women's labor force participation. Given the close connection between investments in human capital and anticipated labor market returns (Becker 1962), we also test whether grammatical gender is associated with larger gender gaps in educational attainment. We then examine the association between grammatical gender and support for traditional gender roles, as measured through a series of questions about gender attitudes included in the World Values Survey.

A limitation of our stylized model is that it is static: it characterizes a psychological phenomenon – the predisposition to partition the world into masculine and feminine domains – in a decentralized equilibrium at a single point in time. In our simple model, high-ability women prefer not to enter the labor force when the proportion of other women doing so is relatively low; they are constrained by their own tastes – by their inclination to frame the world in terms of a dichotomy between what is masculine and what is feminine – not by their culture or by historical institutions. However, an important characteristic of language structure is that it evolves very slowly. For example, scholars believe that Proto-Indo-European (PIE) had separated into distinct offshoots that would evolve into Celtic, Greek, Germanic, Indic, etc. by 2,500 BCE – all of which contained the grammatical gender structure inherited from PIE (Anthony 2007). Hence, much of the variation in grammatical gender observed today reflects changes in language structure that occurred hundreds if not thousands of years ago.

If grammatical gender predisposes people to partition the world into a feminine domain (the home) and a masculine domain (the wider world), this psychological phenomenon has been repeated in every generation for the last several millennia within many linguistic communities. This raises the possibility that what might have begun as a subconscious linguistic nudge has, over time, influenced the evolution of informal and even formal institutions – that a psychological predisposition toward viewing the home as a feminine domain and the outside world as a masculine one has transformed into a constraint on women's access to the world outside the home, limiting their agency and opportunities. Such a gradual diminution of women's autonomy would be consistent with models of bargaining in marriage, which show that changes to one partner's outside options – including those that result from changes in "extrahousehold environmental factors" and those that are the direct consequence of choices made by the couple themselves (e.g. if a woman's decision to leave the labor force reduces her income in the event of divorce) – can lead to a decline in one partner's bargaining power (McElroy 1990, Lundberg and Pollak 2003, Pollak 2019). In the modern era, female labor force participation has also been shown to be influenced by peer effects, particularly the behavior of women in the previous generation, with women more likely to work in adulthood if their high school classmates' mothers' worked (Olivetti, Patacchini and Zenou 2020). Thus, at the level of a society, a linguistic nudge that keeps more women in the home *today* might reduce the likelihood that future generations of women work outside the home, and these traditions may evolve into social norms or even legal restrictions over the course of many generations. The increasing rigidity of traditional gender roles could further limit women's outside options, undermining their autonomy and authority within the home – even if a norm first developed because home was perceived as women's sphere while the world outside was perceived as the male sphere.[13]

We evaluate this potential causal pathway in two ways. First, we test whether country-level variation in grammatical gender is associated with weaker legal protection of women's equality (e.g. through laws that limits women's ability to own property or banning sex discrimination in the workplace). This provides a direct test of the extent to which grammatical gender is associated with gender inequality that goes beyond contemporary attitudes and preferences, and also points to one potential causal mechanism. Second, we test the association between grammatical gender and women's decision-making power directly, with a particular focus on the extent to which women have decision-making autonomy within the home. In our stylized theoretical model, the tendency to partition the world into male and female domains constrains both men and women – if women are being kept out of the labor force, then men are also being

[13] Of course, one might also hypothesize that some aspect of culture that made males dominant also contributed to the emergence of grammatical gender in certain ancestral languages. Since language structures evolve over centuries, it cannot be the case that *present-day* gender attitudes had a causal impact on modern grammatical structures, and most linguists do not believe that language structures evolve to reflect social structures (indeed, many linguists would argue that such a process is not possible – cf. McWhorter (2014) for discussion). However, one cannot completely rule out this mechanism, and 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. We have a relatively good understanding of the process through which grammatical gender was *lost* from certain widely spoken Indo-European languages, and 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). Linguists do not believe that grammatical gender arose as a reflection of historical gender norms (McWhorter 2014). Nevertheless, we do not rule out the possibility of a mechanism from culture to language structure on theoretical grounds; we discuss the approach we take to controlling for prehistoric cultural traits in Section 3.2.

kept out of the domestic sphere. If, however, traditional gender norms have been encoded in formal and informal institutions over time, then we might expect women to have less decision-making power (because of their limited outside options), even in domains that are perceived as feminine. Taken together, this additional set of empirical tests allows us to distinguish exclusively psychological factors that nudge men and women toward separate, autonomous spheres from a story where restrictions on women's ability to participate in economic life outside the home translates, over the course of many generations, into lower overall opportunity and agency.

3 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.[14] 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 previously existing databases cover far fewer languages.

3.1 Building a Grammatical Gender Data Set

Data on the number of speakers of each of the world's languages comes from the *Ethnologue*, a comprehensive database of over 7,000 languages (Lewis, Simons and Fennig, eds., 2016). Combining the *Ethnologue* with information on grammatical gender structure allows us to estimate the fraction of each country's population that speaks a gender language as their mother

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

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 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*; and George Abraham Grierson's eleven-volume *Linguistic Survey of India*, 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 less-documented languages (e.g. Fiji Hindi and Rohingya).[15]

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.[16] Second, we then create an indicator for those gender

[15]Detailed information on the range of sources (including quotes that characterize each language's grammatical gender structure) is provided in the data set, which will be made public as soon as the paper is accepted for publication.

[16]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 rely on sources that characterize grammatical gender structure in one of the following ways: (i) by explicitly stating that a language does or does not have grammatical gender; (ii) by explicitly stating that there are only two genders (male and female);

languages (e.g. Spanish) that have only two noun classes: masculine and feminine.

We successfully classify 4,346 languages which together account for more than 99 percent 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 90 percent of the population in only three countries: Cameroon (89.1 percent), Chad (75.4 percent), and Papua New Guinea (32.0 percent). Figure 1 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.

3.2 Data from the Ethnographic Atlas

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 the machine learning technique, 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

thereby demonstrating that nouns referring to inanimate objects must be assigned to one of those two classes; or (iii) by providing a list of agreement classes that includes masculine and feminine together with a list of words in each class that includes at least one inanimate (or other word that does not refer to living beings with a natural sex or gender).

the significance of the plough is consistent with existing work (Boserup 1970, Alesina et al. 2013). To address concerns about cultural confounds to the greatest extent possible, we include language-level controls for early cultural practices that predict the use of a gender language throughout our analysis.[17]

3.3 Other Sources of Data

Additional cross-country data comes from a number of sources. Data on labor force participation, income, and population come from the World Development Indicators database, while data on the extent to which gender equality is enshrined in the law comes from the World Bank's Women, Business, and the Law data base. We use data on educational attainment from the Barro-Lee Educational Attainment Data Set (Barro and Lee 2013), which contains data on 142 countries. Data on support for traditional gender roles comes from the World Values Survey and is available for 56 (mostly high-income and upper-middle-income) countries. Data on women's decision-making autonomy (available for 67 low-income and lower-middle-income countries) comes from the Demographic and Health Surveys (DHS).[18] Finally, we take several country-level geographic controls (average precipitation and rainfall plus suitability for the plough) from Alesina et al. (2013).

Data for our individual-level analysis comes from two sources. For African countries, we use the nationally-representative Afrobarometer Surveys (Afrobarometer Data 2016). We use Afrobarometer 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. 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

[17]Since language structures pre-date many cultural practices, it is possible that cultural features documented in the Ethnographic Atlas might have been impacted by grammatical gender. All of our results are nearly identical when ethnographic controls are omitted.

[18]The DHS includes nationally representative micro data on women, children, and households. However, because the DHS does not report detailed data on the native languages for all respondents, we use DHS data to construct country-level averages of outcomes of interest.

India Human Development Survey (Desai, Dubey and Vanneman 2015). The IHDS includes data on 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.

4 Traditional Gender Roles

4.1 Cross-Country Analysis

4.1.1 Empirical Strategy

In all of our cross-country analysis, the independent variable of interest is the proportion of a country's population whose native language is a gender language, $GenderLanguageProportion_c$. Our main empirical specification is an OLS regression of the form:

$$Y_c = \alpha + \beta GenderLanguageProportion_c + \delta_{continent} + \lambda X_c + \varepsilon_c \tag{1}$$

where Y_c is the dependent variable in country c, $GenderLanguageProportion_c$ 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.[19] Standard errors are clustered at the level of the most widely spoken language within each country.

In our analysis of adherence to traditional gender roles, the 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 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

[19]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.

force participation minus men's labor force participation).

We also examine support for traditional gender roles using data on attitudes 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 traditional gender roles. Since we are considering attitudes rather than behaviors, we do not report gender differences; we instead test whether grammatical gender predicts support for traditional gender roles among both men and women.

4.1.2 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: 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. 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 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.

In Online Appendix Table A1, we present an alternative specification that captures potential variation in treatment intensity resulting from weaker vs. stronger forms of grammatical gender. Languages with only two noun classes – masculine and feminine – might be expected to create a stronger subconscious inclination to see the world in terms of a gender dichotomy, as Whorf hypothesized. To test this, we construct a *grammatical gender intensity index* that is the proportion of a country's population that speaks a strong gender language (with only two noun classes, masculine and feminine) plus 0.5 times the proportion of the population that speaks a weak gender language with more than two genders (e.g. a language such as German, which has masculine, feminine, and neuter noun classes).[20] The results reported in Online Appendix Table A1 again suggest a robust negative association between our continuous measure of grammatical gender intensity and women's labor force participation, with coefficients that are similar in magnitude to those reported in Table 1.[21]

In Figure 2, 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.[22] 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

[20]Our continuous measure is similar to the *gender intensity index* proposed by Gay et al. (2018). Because they focus on a sample of countries that can be characterized by a single dominant or official language, they use an index that captures variation in treatment intensity across languages, but not cross-country variation in treatment intensity that results from ethnolinguistic diversity. Their approach is not well-suited to analysis of data from countries with high levels of linguistic diversity such as those in Sub-Saharan Africa (Easterly and Levine 1997).

[21]In Online Appendix Table A2, we include both $GenderLanguageProportion_c$, our main independent variable, and a separate variable indicating the proportion of the population whose native language is a strong gender language with only two noun classes. Results suggest that much of the association between gender languages and women's labor force participation is driven by languages with only two noun genders, but the use of any gender language may help to explain the observed variation in the gender gap in labor force participation. We also report a range of additional 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 A3); when we drop each of the major world languages of Arabic, English, and Spanish (Online Appendix Table A4; and when we include of a set of "bad controls" that could have been impacted by grammatical gender (Online Appendix Table A5. As is well known, including such controls could bias the coefficient of interest, making it impossible to interpret (Angrist and Pischke 2008). Nevertheless, results are broadly similar when we control for log GDP per capita, population, major world religions, and an indicator for post-Communist regimes.

[22]Data on female labor force participation is available from the World Bank for every year since 1990.

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.

4.1.3 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 earnings 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, 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 attainment 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): grammatical gender is associated with a 0.6 year increase in the gender gap in years of schooling (p-value 0.026). We also observe a negative and statistically significant relationship between women's educational attainment and our continuous measure of exposure to grammatical gender (Online Appendix Table A6).[23]

[23] When we include separate independent variables for the proportion of a country's population that speaks any

The Barro-Lee Educational Attainment Dataset provides estimates of adult educational attainment at five-year intervals from 1950 through 2010. In Figure 3, 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 2021). Whether one considers the level of female educational attainment or the gender gap in schooling, there is no indication that the cross-country association with grammatical gender has diminished over time.

4.1.4 Attitudes about Traditional Gender Roles

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 measuring support for traditional gender roles. In Figure 4, we plot the cross-country relationship between responses to each of these questions and the proportion of a country's population whose native language is a gender language. The prevalence of 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 estimates suggests that grammatical gender is associated with greater support for traditional gender roles. In Columns 3 through 6 of Table 3, we show that there is a negative association between the country-level prevalence of grammatical gender and traditional gender attitudes among both women (Columns 3 and 4) and men (Columns 5 and 6). Though the coefficient is slightly larger

gender language and the proportion that speaks a strong gender language with only two genders, we find that variation in the level of women's educational attainment is primarily explained by variation in the prevalence of strong (dichotomous) gender languages, but cross-country variation in the gender gap in educational attainment is related to both measures of grammatical gender prevalence (Online Appendix Table A7). We also report a range of additional robustness checks in Online Appendix Tables A10 through A12. Results are similar if we use the rate of (or the gender gap in) primary school completion as the dependent variable (Online Appendix Table A8). We do not observe a statistically significant association between grammatical gender and women's likelihood of completing secondary school (Online Appendix Table A9). Results are qualitatively similar if we calculate the gender gap in educational attainment as a ratio rather than a linear difference (Online Appendix Table A10), omit countries where the most widely spoken language is English, Spanish, or Arabic (Online Appendix Table A11), or include a range of "bad controls" including current GDP per capita (Online Appendix Table A12).

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 support for traditional gender roles among both men and women.

4.1.5 Robust Inference

In this section, we discuss two potential concerns with the estimation approach used in our cross-country analysis. First, as discussed above, we were unable to classify the gender structure of some languages. We therefore 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. We then 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.

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 1. 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:

$$CI_\alpha = [\hat{\beta}^l - \bar{C} \cdot SE(\hat{\beta}^l), \hat{\beta}^u + \bar{C} \cdot SE(\hat{\beta}^u)] \tag{2}$$

where $\bar{C}$ satisfies

$$CDF\left(\bar{C} + \frac{\hat{\Delta}}{\max(SE(\hat{\beta}^l), SE(\hat{\beta}^u))}\right) - CDF(-\bar{C}) = \alpha \tag{3}$$

for the CDF of Student's t-distribution with the appropriate number of degrees of freedom.[24] 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 previously available in the World Atlas of Language Structures, Figure 5 shows that the Imbens-Manski confidence intervals would be substantially wider and would always include zero, because grammatical gender data had previously been available only for 64 percent (rather than 99 percent) of the world's population. Thus, our data set allows for more robust inference than had previously been possible.

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

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 families.[25]

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.

This approach defines a set of 203 clusters, 69 of which have grammatical gender.[26] Having assigned all the languages to clusters in this manner, we conduct a permutation test by randomly generating counterfactual allocations of gender structure that would be possible while holding fixed the structure of the treatment variation across the language tree and the number of clusters "treated" with grammatical gender (69 of 203). We use each such counterfactual assignment of treatments to create an associated country-level measure of grammatical gender

[25]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.

[26]Roughly one quarter of these clusters are at the highest, language family level; one quarter are individual languages; and the remaining half (including Nilo-Saharan and Indo-European languages) are clustered at a level in between.

(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.

4.2 Within-Country Analysis

4.2.1 Empirical Strategy

Next, we explore the relationship between gender languages and women's education and 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. In these countries, both gender and non-gender languages are indigenous — in contrast to, for example, several countries in South America where non-gender indigenous languages and a gender colonial language are both widely spoken.[27] 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

[27] For example, Maasai and Luo, two languages from the Nilo-Saharan language family that are widely spoken in Kenya, differ in their grammatical gender structure: Maasai is a gender language, while Luo is not.

allow us to examine the relationship between grammatical gender and women's outcomes while holding much of the cultural and institutional context constant.

Our individual-level analysis parallels our cross-country analysis, using data from Afrobarometer surveys and the India Human Development Survey. 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:

$$Y_i = \alpha + \beta Gender_i + \gamma Z_i + \varepsilon_i \tag{4}$$

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 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:

$$Y_i = \alpha + \beta Gender_i + \zeta Female_i + \mu Gender \times Female_i + \gamma Z_i + \varepsilon_i \tag{5}$$

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.

4.2.2 Results

We summarize our regression results in Figure 6 (regression results are presented in Online Appendix Tables A15 through A26). 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 6), 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 6), 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 in two distinct cultural contexts, even when restricting attention to indigenous languages that differ in terms of their grammatical gender structure.

5 Other Measures of Women's Equality

Our analysis thus far has focused on traditional gender roles, influenced by Whorf's argument that the practice of partioning the set of nouns into masculine and feminine might create a subconscious predisposition to partition other aspects of the human experience in a similar manner. As suggested by Becker (1981), traditional gender roles – where the domestic sphere is perceived as a feminine domain while the labor market is perceived as a man's world – represent one of the most widely accepted examples of such a sex-based partition of domains. Our empirical results suggest a strong association between the use of gender languages and adherence to traditional gender roles, both across and within countries. We now turn from outcomes relating to partitions of the world into essentially equal or comparable masculine and feminine domains, to measures of women's overall agency or agency within the feminine domain. We consider two examples of outcomes in this category. First, we examine the association between grammatical gender and laws guaranteeing equal rights and freedoms for men and women. Second, we consider the relationship between grammatical gender and decision-making within the household – with a particular focus on domains (e.g. a woman's own health, the home) that might be considered part of the feminine sphere within the context of traditional gender roles.

5.1 Is Women's Equality Enshrined in the Law?

Data on the extent to which women and men are treated equally under the law comes from the World Bank's Women, Business, and the Law (WBL) database. WBL is constructed by interviewing in-country legal experts to determine whether a country treats men and women equally under the law and protects women from sex/gender discrimination. For example, the database includes information on whether men and women have an equal right to inherit property, whether women can initiate a divorce, whether the law bans sex discrimination in the workplace and mandates equal pay for equal work, and whether women are allowed to open a bank account or take out a loan in the same way as a man. We construct an index of 29 WBL measures related to equality under the law, and then regress this index on our country-level measure of the prevalence of grammatical gender. Results are reported in Table 6. After controlling for continent fixed effects, we find a strong negative relationship between the prevalence of grammatical gender and legal support for gender equality: the coefficient estimate suggests that countries where the entire population speaks a gender native language will have, on average, three fewer laws protecting women's equality than those where no one speaks a gender native language.

In Panels B, C, and D of Table 6, we disaggregate the WBL data into three sub-indices: one capturing family law and rights related to asset ownership, one capturing laws related to employment and participation in credit markets, and one capturing freedom of movement. One might expect that a culture with stronger traditional gender norms might have fewer laws protecting women from workplace discrimination (since women are expected to be in the home) – and that is exactly what we find. However, we also find evidence that grammatical gender predicts a lack of equalty in family law, suggesting that women have weaker rights within the household. This appears inconsistent with the idea of separate spheres, since women might be expected to have at least as much authority within the home in societies where home is perceived as a feminine domain. We do not find evidence of a statistically significant relationship between grammatical gender and laws related to women's freedom of movement.

These results suggest that grammatical gender is not simply an indicator of a difference in subconscious determinants of individual preferences: countries where gender languages are more prevalent have fewer laws protecting women's equality and prohibiting sex discrimination. These differences in legal environment impact women irrespective of their mother tongue, and may continue to constrain women's opportunities after support for traditional gender norms wanes. That said, these results are also policy relevant: policymakers have many reasons to be cautious about attempting to influence the evolution of language structure (given the culturally fraught and sometimes quixotic history of such efforts), but no linguistic heritage is threatened by modifications to the legal code that would outlaw discrimination and mandate the equal treatment of men and women.

5.2 Women's Decision-Making Autonomy in the Domestic Sphere

In our final piece of analysis, we examine the association between grammatical gender and decision-making power within the household. To do this, we use Demographic and Health Survey (DHS) data from 67 low-income and middle-income countries. Married female respondents are asked who makes decisions about the woman's health, who makes decisions about purchasing household items, and who makes decisions about visits to family. If grammatical gender is associated with a tendency to partition the world in to distinct masculine and feminine domains, we might expect women to have greater decision-making autonomy within their own sphere – for instance, when it comes to their own health, or to decisions about the purchase of small household items. If, on the other hand, support for traditional gender norms has limited women's bargaining power and outside options over many generations, we might expect women in countries where gender languages are prevalent to have less decision-making power, even within seemingly female domains.

To test this, we construct indicators for independent decision-making by women and men. We define an indicator for women's decision-making autonomy that is equal to one if a DHS respondent indicates that she makes decisions in a given domain (about her own health, purchasing household items, and visiting family) independently. We define an indicator for men's decision-making power that is equal to one if a woman says that her husband makes decisions in these domains on his own (we use "power" rather than "autonomy" since these are mainly decisions related to the feminine domain).[28] We then average our three measures of women's autonomy (decisions about her own health, decisions about household purchases, and decisions about visits to family) to construct an overall index of women's decision-making agency, and do the same with our three measures of men's decision-making power.

Results are reported in Table 6. After controlling for continent fixed effects and geography and ethnography controls, our findings suggest that grammatical gender predicts reduced female decision-making autonomy, even in domains likely to be perceived as female. Coefficient estimates suggest that grammatical gender is associated with an eight percentage point reduction in the likelihood that women make decisions on their own (p-value 0.04) and a 14 percentage point increase in the likelihood that men make household decisions unilaterally (p-value 0.026). We also see several other statistically significant patterns: a decrease in the likelihood that a woman can make unilateral decisions about household purchases, an increase in the likelihood that a man can make unilateral decisions about household purchases, and a decrease in the likelihood that a woman can make unilateral decisions about visiting family. Thus, grammatical gender is not associated with an increased tendency to partition the world into a "wider world" where men dominate and a feminine domestic sphere where women are in charge. Instead, our results suggest that grammatical gender is associated with reduced female autonomy, even within the home.

[28]These two dummies do not sum to one: a woman could also indicate that she and her husband make decisions together, that one of them makes decisions together with a third party, or that a third party makes decisions in that domain without input from the couple.

6 Conclusion

Using a new data set on the grammatical gender structure of more than 4,000 languages, we document a robust association between gender languages and traditional gender roles. 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. We also show that grammatical gender 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.

The caveat, of course, is that these associations are correlations, and not necessarily causal relationships. In most cases, whether a language has retained grammatical gender is driven by idiosyncrasies 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 – and not because of changes in gender norms in pre-Norman England (McWhorter 2005, Kastovsky 1999). Nevertheless, gender languages are not randomly assigned, and the observed correlations might be driven by some unobserved causal factor that is correlated with both language and support for traditional gender roles.

Because grammatical structures evolve over many centuries, modern gender norms could not explain the observed empirical relationship between grammatical gender and women's equality. Reverse causality is ruled out. Instead, any alternative causal mechanism must involve some pre-modern cultural characteristic that could have shaped both linguistic structure and gender norms.[29] We address this issue to the extent possible by combining our linguistic database with the Ethnographic Atlas (Murdock 1967) and using simple machine learning techniques to

[29]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.

identify the pre-modern cultural traits that predicted use of a gender language; these variables are included as controls throughout our analysis. We also control for religious affiliation in our within-country specifications and as a robustness check in our cross-country analysis.[30] However, we cannot fully rule out the possibility that cultural factors shaped both grammatical structures and gender norms. As in all studies of history and culture, it is not possible to run experiments and the number of independent observations is fairly small; some measure of caution about straightforward causal interpretation is therefore warranted.

Nevertheless, 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, we also find evidence that grammatical gender is associated with legal structures that limit women's equality, and with weaker decision-making power within the household. Hence, grammatical gender should not be seen through a purely behavioral economic (or psychological) lens. Our results suggest a subtle linguistic factor that has nudged women and men toward traditional gender roles over many centuries gradually contributed to the evolution of institutions that continue to limit women's equality in the present.

[30] In earlier versions of this paper, we also followed the coefficient stability approach suggested by Altonji, Elder and Taber (2005) and further refined by Oster (2017). Results suggest that our findings are sufficiently robust to the inclusion of controls to suggest that the observed association is unlikely to be explained by unobservables (Jakiela and Ozier 2019).

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.

Akerlof, George A. and Rachel E. Kranton, "Economics and Identity," *Quarterly Journal of Economics*, 2000, *115* (3), 715–753.

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.

Anthony, David W., *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World*, Princeton University Press, 2007.

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.

Barber, Elizabeth Wayland, *Women's Work: The First 20,000 Years*, W. W. Norton and Comany, 1994.

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.

Beblo, Miriam, Luise Görges, and Eva Markowsky, "Gender Matters in Language and Economic Behaviour: Can we Measure a Causal Cognition Effect of Speaking?," *Labour Economics*, 2020, *65*, 101850.

Becker, Gary S, "Investment in Human Capital: A Theoretical Analysis," *Journal of Political Economy*, 1962, *70* (5), 9–49.

—— , *A Treatise on the Family*, Harvard University Press, 1981.

—— , "Human Capital, Effort, and the Sexual Division of Labor," *Journal of Labor Economics*, 1985, *3* (1), S33–S58.

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.

Brown, Judith K., "A Note on the Division of Labor by Sex," *American Anthropologist*, 1970, *72* (5), 1073–1078.

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.

Easterly, William and Ross Levine, "Africa's Growth Tragedy: Policies and Ethnic Divisions," *Quarterly Journal of Economics*, 1997, *112* (4), 1203–1250.

Evans, David K., Maryam Akmal, and Pamela Jakiela, "Gender Gaps in Education: The Long View," *IZA Journal of Development and Migration*, 2021, *12* (1).

Fabb, Nigel, "Linguistic theory, linguistic diversity and Whorfian economics," in "The Palgrave handbook of economics and language," Springer, 2016, pp. 17–60.

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*, 2018, *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 the-Marathi 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 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.

Jakiela, Pamela and Owen Ozier, "Gendered Language," Center for Global Development Working Paper 500 2019.

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].

Lundberg, Shelly and Robert A. Pollak, "Separate Spheres Bargaining and the Marriage Market," *Journal of Political Economy*, 1993, *101* (6), 988–1010.

___ **and** ___ , "Efficiency in Marriage," *Review of Economics of the Household*, 2003, *1* (3), 153–167.

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.

McElroy, Marjorie B., "The Empirical Content of Nash-Bargained Household Behavior," *Journal of Human Resources*, 1990, pp. 559–583.

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.

Olivetti, Claudia, Eleonora Patacchini, and Ives Zenou, "Mothers, Peers, and Gender-Role Identity," *Journal of the European Economic Association*, 2020, *18* (1), 266–301.

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.

Pollak, Robert A, "How Bargaining in Marriage Drives Marriage Market Equilibrium," *Journal of Labor Economics*, 2019, *37* (1), 297–321.

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.

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.

Tabellini, Guido, "Institutions and Culture," *Journal of the European Economic association*, 2008, *6* (2-3), 255–294.

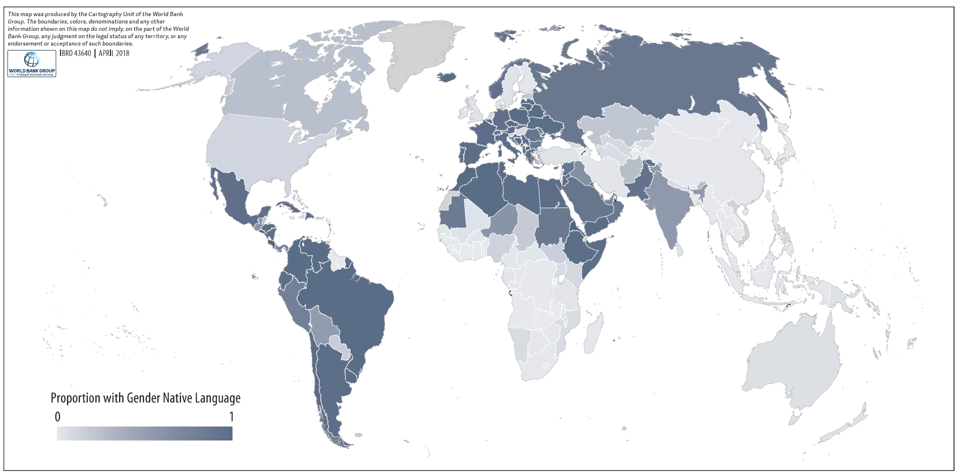
Trigger, Bruce E., *Understanding Early Civilizations*, Cambridge University Press, 2003.

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.

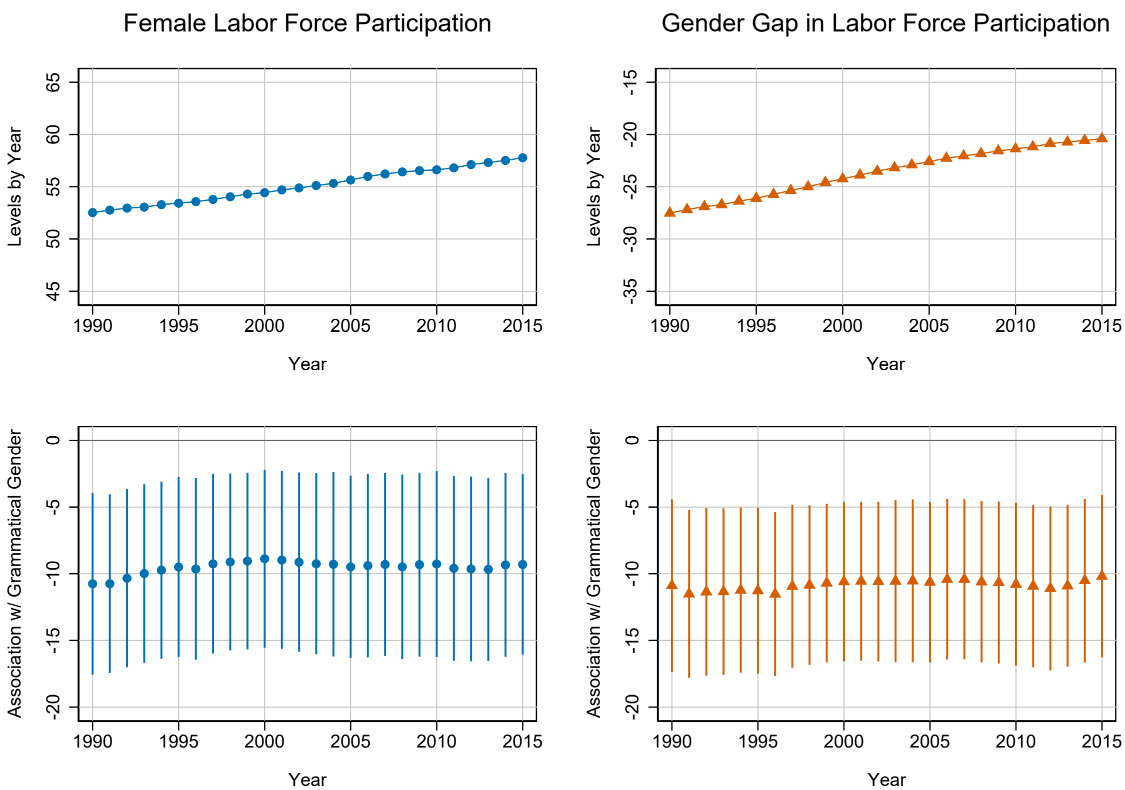
Winawer, Jonathan, Nathan Witthoft, Michael C Frank, Lisa Wu, Alex R Wade, and Lera Boroditsky, "Russian blues reveal effects of language on color discrimination," *Proceedings of the National Academy of Sciences*, 2007, *104* (19), 7780–7785.

Figure 1: The Distribution of Gender Languages



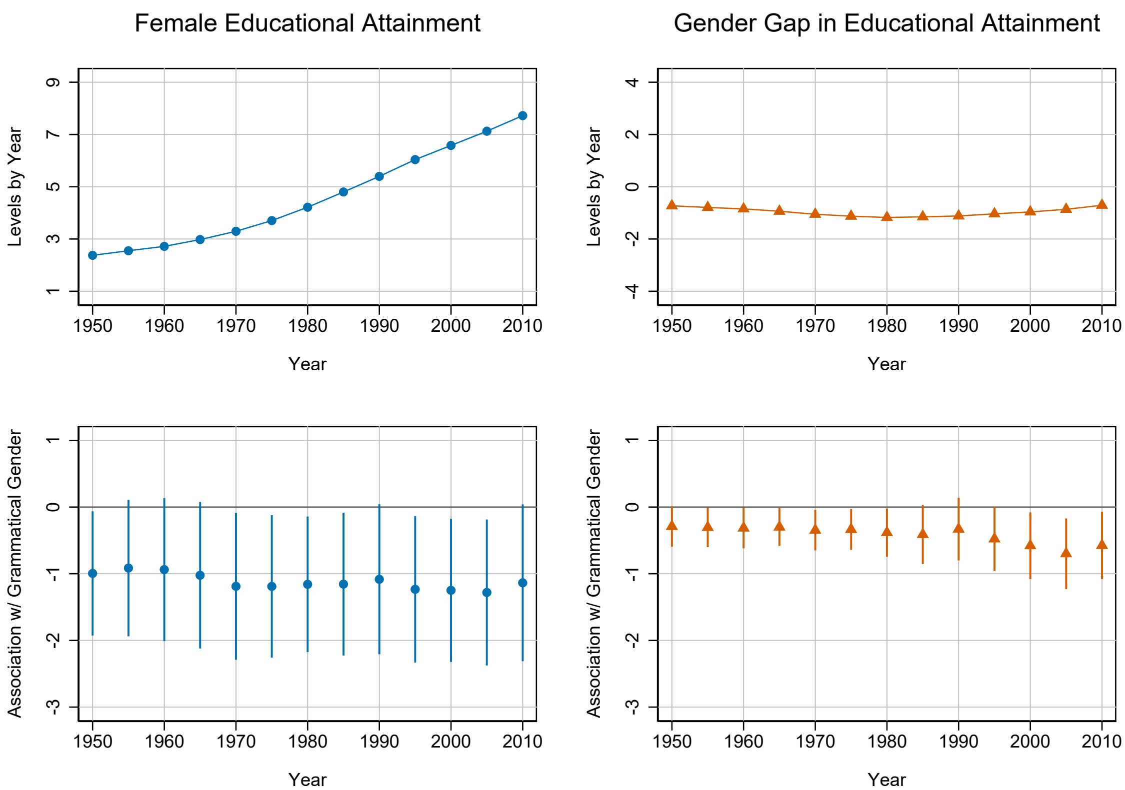
The figure shows the percentage of the native speakers in each country whose native language is a gender language (i.e. the fraction of *Ethnologue* native speakers whose native language uses a system of grammatical gender). The figure assumes that missing data (on 0.8 percent of all native speakers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

Figure 2: Labor Force Participation and Grammatical Gender Over Time



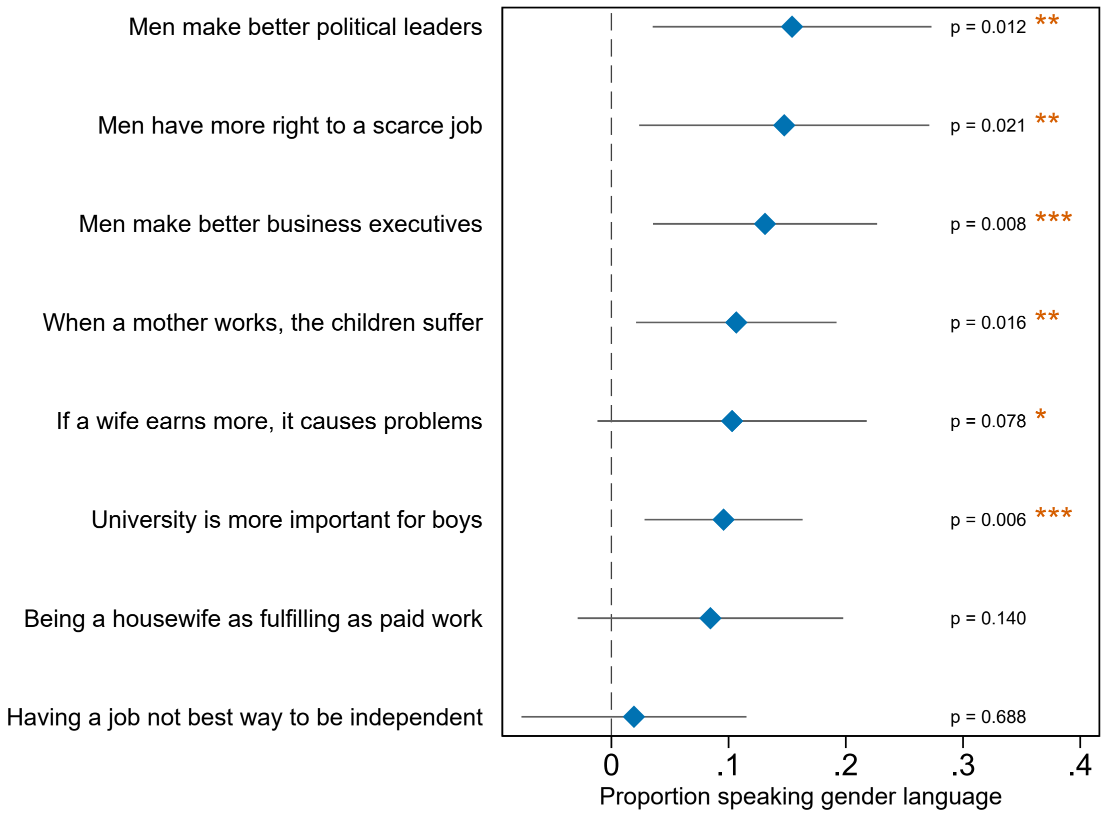
Data on labor force participation is available from the World Bank's World Development Indicators database for years 1990 through 2017. The gender gap in labor force participation is measured as the difference between female and male labor force participation. Lower panels report OLS coefficients and 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 3: 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 4: Cross-Country Variation in Gender Attitudes



The figure summarizes the results from a series of regressions of (country-level averages of) responses to World Values Survey (WVS) questions on the proportion of a country's population whose native language is a gender language. We present the results for all eight WVS questions related to gender attitudes. Responses to all eight questions are coded so that the answer most consistent with traditional gender norms (involving separate roles for men and women) is equal to 1 and the response most consistent with gender equality is equal to 0. Each regression is estimated via OLS and includes continent fixed effects 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 5: Comparison of Unadjusted and Manski-Imbens Confidence Intervals, New vs. WALS-Only Data

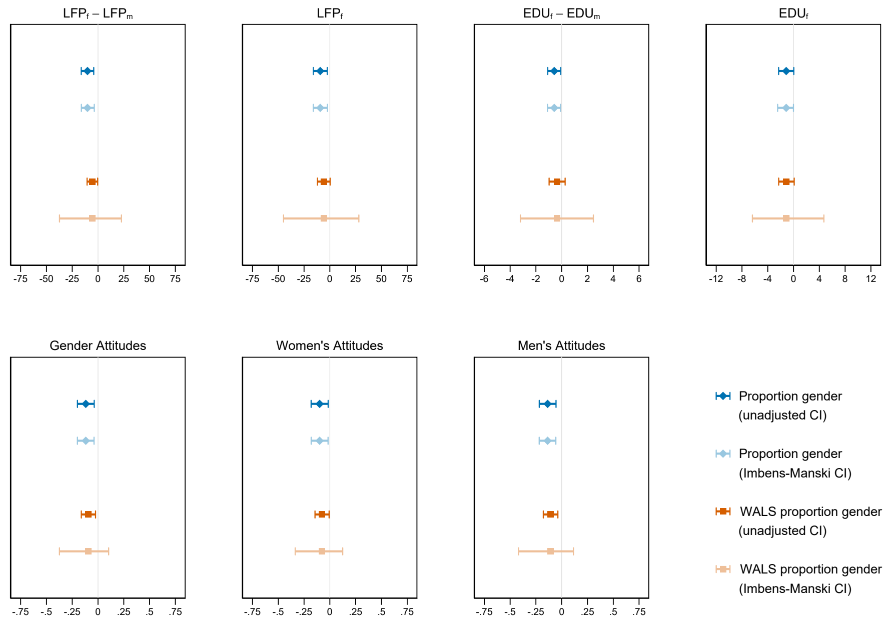
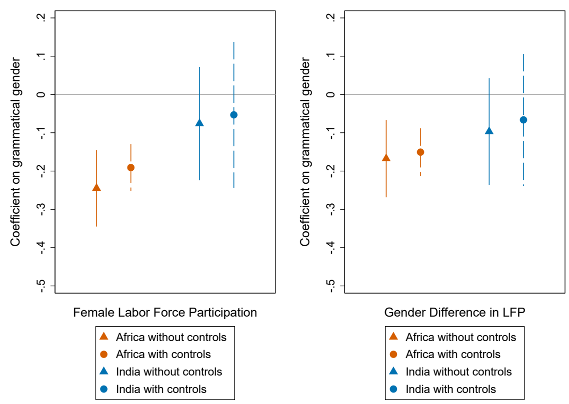


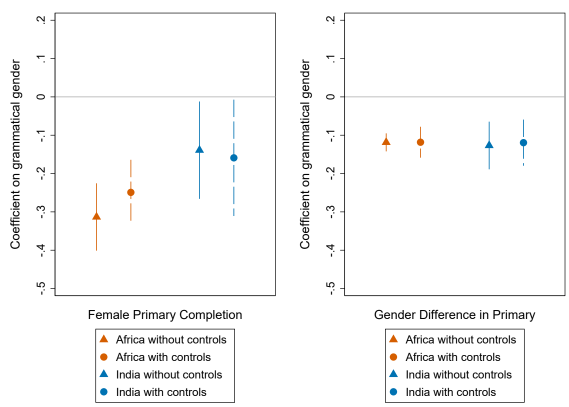
Figure illustrates point estimates (the diamond or square) and 95-percent confidence intervals for cross-country regressions of seven different outcomes on the proportion of a country's population that speaks a gender language. The upper intervals (shades of blue) use our new comprehensive data. The lower intervals (shades of orange) use only grammatical gender data from the World Atlas of Language Structures (WALS). For each pair of intervals, the standard OLS cluster-robust confidence interval (treating as ignorable any languages whose gender status is not coded in the relevant data) is presented in the darker shade above, while the worst-case bounds for missing data are presented via the Imbens-Manski confidence interval in the lighter shade below. For each outcome, the naïve, "unadjusted," confidence interval comes from the associated regression (with controls) in Table 1, 2, or 3. 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. 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. Units for labor force participation are percentage points; units for education are years of schooling; attitudes (lower panels) are on a scale from zero to one. The procedures for calculating Imbens-Manski confidence intervals is described in section 4.1.5.

Figure 6: Within-Country Variation in Grammatical Gender

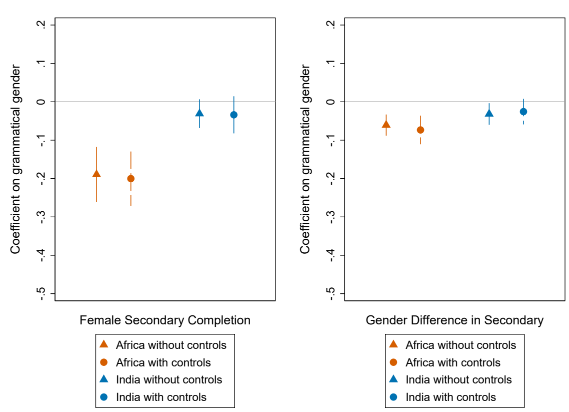
Panel A: Labor Force Participation



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 Support for Traditional Gender Roles

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]
Continent Fixed Effects	No	Yes	No	Yes	No	Yes
Geography/Ethnography Controls	No	Yes	No	Yes	No	Yes
Observations	56	56	56	56	56	56
R^2	0.01	0.78	0.00	0.74	0.02	0.79

Robust standard errors clustered by most widely spoken language in all specifications. Our index of Gender Attitudes is constructed by taking the first principal component of the eight World Values Survey questions measuring support for traditional gender norms (described in Figure 4) at the individual level, and then calculating the average of this index within a country. Numbers closer to one 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 4.1.5. 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 4.1.5. 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.

Table 6: Grammatical Gender and Women's Equality in the Law

	OLS (1)	OLS (2)
Panel A. Dependent Variable: Gender Equality Enshrined in Law		
Proportion speaking gender language	-0.34	-2.95
	(1.38)	(0.97)
	[0.803]	[0.003]
Panel B. Dependent Variable: Gender Equality in Family/Inheritance Law		
Proportion speaking gender language	-0.36	-1.18
	(0.64)	(0.48)
	[0.571]	[0.014]
Panel C. Dependent Variable: Gender Equality in Contract/Employment Law		
Proportion speaking gender language	-0.00	-1.54
	(0.61)	(0.51)
	[0.998]	[0.003]
Panel D. Dependent Variable: Gender Equality in Freedom of Movement Law		
Proportion speaking gender language	0.02	-0.23
	(0.23)	(0.20)
	[0.932]	[0.237]
Continent FEs	No	Yes
Geography/Ethnography Controls	No	Yes
Observations	174	174

Robust standard errors clustered by most widely spoken language in all specifications. The outcome variable in Panel A is an index of 29 measures of legal equality (as of 2015) included in the World Bank's Women, Business, and the Law (WBL) database (we omit six questions related to the way that benefits are administered). The outcome variable in Panel B is an index of 10 WBL measures of legal equality related to family structure family structure (e.g. can a woman be the head of a household) and women's inheritance and property ownership rights. The outcome variable in Panel C is an index of 15 WBL measures of legal equality related to employment, sex/gender/pregnancy discrimination in the workplace, and access to credit and business opportunities. The outcome variable in Panel D is an index of four WBL measures of legal equality related to freedom of movement (e.g. can a woman obtain a passport without permission from a male family member). All indices are constructed so that a regression coefficient of one would indicate that, all else equal, countries with 100 percent of the population speaking a gender native language have an average of one fewer law protecting women's equality relative to countries where zero percent of the population speaks a gender native language. 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 7: Women's Decision-Making Autonomy

	Women's Autonomy		Men's Power	
	OLS	OLS	OLS	OLS
	(1)	(2)	(3)	(4)
Panel A. Dependent Variable: Decision-Making Index				
Proportion speaking gender language	-0.00	-0.08	-0.09	0.14
	(0.05)	(0.04)	(0.05)	(0.06)
	[0.988]	[0.040]	[0.092]	[0.026]
Panel B. Dependent Variable: Decisions About Women's Health				
Proportion speaking gender language	0.07	-0.07	-0.12	0.17
	(0.07)	(0.07)	(0.06)	(0.07)
	[0.325]	[0.349]	[0.051]	[0.021]
Panel C. Dependent Variable: Decisions About Purchasing Household Items				
Proportion speaking gender language	-0.05	-0.07	-0.07	0.16
	(0.03)	(0.03)	(0.05)	(0.06)
	[0.159]	[0.028]	[0.152]	[0.010]
Panel D. Dependent Variable: Decisions About Visiting Family				
Proportion speaking gender language	-0.02	-0.10	-0.11	0.08
	(0.04)	(0.03)	(0.05)	(0.06)
	[0.628]	[0.007]	[0.037]	[0.217]
Continent FEs	No	Yes	Yes	Yes
Geography/Ethnographic Controls	No	Yes	No	Yes
Observations	67	67	67	67

Robust standard errors clustered by most widely spoken language in all specifications. The outcome variable in Panel A is an index of three DHS questions (that were asked to women) about who makes household decisions (about the woman's health, purchasing household items, and visiting family). The dummy for Women's Autonomy used in Columns 1 and 2 indicates that the woman reported making decisions on her own. The dummy for Men's Power used in Columns 3 and 4 indicates that the woman reported that her husband made decisions (for example, about the woman's health) on *his* own, without input from the wife or anyone else. Women could also report that she and her husband made decisions together, that either of them made decisions together with a third party (e.g. a mother or mother-in-law), or that someone else made decisions. 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).

A Online Appendix: not for print publication

Table A1: Cross-Country OLS Regressions of Labor Force Participation on a Continuous Measure of Exposure to Grammatical Gender

Dependent variable:	LFP_f		LFP_f - LFP_m	
Specification:	OLS (1)	OLS (2)	OLS (3)	OLS (4)
Grammatical gender intensity index	-14.50	-10.94	-17.52	-13.15
	(4.24)	(3.71)	(3.30)	(3.01)
	[$p < 0.001$]	[0.004]	[$p < 0.001$]	[$p < 0.001$]
Continent Fixed Effects	No	Yes	No	Yes
Geography/Ethnography Controls	No	Yes	No	Yes
Observations	178	178	178	178
R^2	0.12	0.37	0.19	0.54

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. The Grammatical Gender Intensity Index is a weighted average of the use of grammatical gender within a country, calculated as the proportion of a country's population whose mother tongue uses strong grammatical gender system with only two genders (masculine and feminine) plus 0.5 times times the proportion of the population whose mother tongue uses a weaker grammatical gender system with three or more genders (e.g. masculine and feminine plus neuter, as in German). 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 — 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	6.38	6.38	-3.71
	(3.13)	(5.69)	(1.83)	(5.24)
	[0.187]	[0.276]	[$p < 0.001$]	[0.481]
Proportion speaking strong (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. The proportion of a country's population whose mother tongue uses strong (dichotomous) grammatical gender system is the proportion whose mother tongue uses only two genders (masculine and feminine). 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 Labor Force Participation Gender 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 A4: Cross-Country Regressions of LFP — Dropping Major World Languages

Dependent variable:	LFP_f			$\text{LFP}_f - \text{LFP}_m$		
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	-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. $\text{LFP}_f - \text{LFP}_m$ 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 A5: 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 A6: Cross-Country OLS Regressions of Educational Attainment on a Continuous Measure of Exposure to Grammatical Gender

Dependent variable:	EDU_f		EDU_f - EDU_m	
Specification:	OLS (1)	OLS (2)	OLS (3)	OLS (4)
Grammatical gender intensity index	0.81	-1.89	0.21	-0.70
	(0.75)	(0.64)	(0.31)	(0.31)
	[0.281]	[0.004]	[0.509]	[0.027]
Continent Fixed Effects	No	Yes	No	Yes
Geography/Ethnography Controls	No	Yes	No	Yes
Observations	142	142	142	142
R^2	0.01	0.65	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. The Grammatical Gender Intensity Index is a weighted average of the use of grammatical gender within a country, calculated as the proportion of a country's population whose mother tongue uses strong grammatical gender system with only two genders (masculine and feminine) plus 0.5 times times the proportion of the population whose mother tongue uses a weaker grammatical gender system with three or more genders (e.g. masculine and feminine plus neuter, as in German). 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 A7: Cross-Country Regressions of Educational Attainment — Weak vs. Strong Gender Categories

Dependent variable:	EDU_f		EDU_f - EDU_m	
Specification:	OLS (1)	OLS (2)	OLS (3)	OLS (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. The proportion of a country's population whose mother tongue uses strong (dichotomous) grammatical gender system is the proportion whose mother tongue uses only two genders (masculine and feminine). 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 A8: 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 A9: 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 A10: 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 A11: 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 A12: 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 A13: Cross-Country OLS Regressions of Support for Traditional Gender Norms on a Continuous Measure of Exposure to Grammatical Gender

Dependent variable:	GENDER ATTITUDES		WOMEN'S ATTITUDES		MEN'S ATTITUDES	
Specification:	OLS	OLS	OLS	OLS	OLS	OLS
	(1)	(2)	(3)	(4)	(5)	(6)
Grammatical gender intensity index	-0.07	-0.14	-0.05	-0.12	-0.07	-0.16
	(0.07)	(0.04)	(0.06)	(0.04)	(0.08)	(0.05)
	[0.358]	[0.002]	[0.417]	[0.005]	[0.357]	[$p < 0.001$]
Continent Fixed Effects	No	Yes	No	Yes	No	Yes
Geography/Ethnography Controls	No	Yes	No	Yes	No	Yes
Observations	56	56	56	56	56	56
R^2	0.05	0.78	0.03	0.74	0.05	0.79

Robust standard errors clustered by most widely spoken language in all specifications. Our index of GENDER ATTITUDES is constructed by taking the first principal component of the eight World Values Survey questions measuring support for traditional gender norms (described in Figure 4) at the individual level, and then calculating the average of this index within a country. Numbers closer to one 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 A14: Cross-Country OLS Regressions of Support for Traditional Gender Norms — Weak vs. Strong Gender Categories

Dependent variable:	Gender Attitudes		Women's Attitudes		Men's Attitudes	
Specification:	OLS	OLS	OLS	OLS	OLS	OLS
	(1)	(2)	(3)	(4)	(5)	(6)
Proportion speaking (any) gender language	0.06	0.07	0.07	-0.08	0.05	-0.12
	(0.05)	(0.08)	(0.04)	(0.08)	(0.05)	(0.08)
	[0.190]	[0.218]	[0.130]	[0.330]	[0.334]	[0.118]
Proportion speaking strong (dichotomous) gender language	-0.13	-0.04	-0.12	-0.04	-0.12	-0.03
	(0.07)	(0.10)	(0.07)	(0.10)	(0.08)	(0.10)
	[0.097]	[0.685]	[0.079]	[0.677]	[0.151]	[0.765]
Continent Fixed Effects	No	Yes	No	Yes	No	Yes
Geography/Ethnography Controls	No	Yes	No	Yes	No	Yes
Observations	56	56	56	56	56	56
R^2	0.11	0.79	0.11	0.74	0.10	0.79

Robust standard errors clustered by most widely spoken language in all specifications. Our index of Gender Attitudes is constructed by taking the first principal component of the eight World Values Survey questions measuring support for traditional gender norms (described in Figure 4) at the individual level, and then calculating the average of this index within a country. Numbers closer to one indicate more support for gender equality. The proportion of a country's population whose mother tongue uses strong (dichotomous) grammatical gender system is the proportion whose mother tongue uses only two genders (masculine and feminine). 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 A15: 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 A16: 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 A17: 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 (0.04) [$p < 0.001$]	-0.25 (0.04) [$p < 0.001$]
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 A18: 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 A19: 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 A20: 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 A21: 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 A22: 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 A23: 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 A24: 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 A25: 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 A26: 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).

B Theoretical Model

Most existing work examining the empirical relationship between grammatical gender and women's involvement in economic life has not formally specified the potential causal pathway.[31] 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 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.[32]

B.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

[31] Beblo, Görges and Markowsky (2020) is a notable recent exception.

[32] 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.

boys in the population, girls represent proportion

$$P^*_{girls} = \frac{1 - F_\gamma(\gamma^*)}{2 - F_\beta(\beta^*) - F_\gamma(\gamma^*)} \tag{B1}$$

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

There are three possible equilibria: the M-equilibrium (school is a masculine domain), the N-equilibrium (school is not a gendered domain), and the F-equilibrium (school is a feminine domain). Each equilibrium is associated with a cost structure. For example, because school is a masculine domain in the M-equilibrium, girls face a psychic cost when they choose to go to school (but boys do not). The M-equilibrium only exists if the proportion of girls (as a share of all students) that results from that cost structure is below λ.

If it exists, the N-equilibrium is identical to the case discussed above (in the absence of grammatical gender). Since school is not seen as a gendered domain, boys and girls attend whenever the net return is greater than zero – i.e. whenever $\gamma_i \geq \gamma^N = \gamma^*$ (where γ^* is the solution to $R_g(\gamma^*) = 0$, as discussed in Section 3.1) or $\beta_i \geq \beta^N = \beta^*$. Hence, the proportion of students who are female is equal to

$$P^N_{girls} = \frac{1 - F_\gamma(\gamma^*)}{2 - F_\beta(\beta^*) - F_\gamma(\gamma^*)}. \tag{B2}$$

The N-equilibrium exists if and only if $\lambda \leq P^N_{girls} \leq 1 - \lambda$. In the N-equilibrium, the proportion of children attending school is $1 - [F_\beta(\beta^*) + F_\gamma(\gamma^*)]/2$.

School is considered a masculine domain if the proportion of students who are female is below λ. In this case, girls will only attend school if $R_g(\gamma_i) > \phi$. Define γ^M as the solution to $R_g(\gamma_i) - \phi = 0$. Since the return to education is increasing in ability, $\gamma^M > \gamma^N$. Girls with $\gamma_i \geq \gamma^M$ will attend school whether or not school is perceived as a masculine domain, but those with $\gamma_i \in (\gamma^N, \gamma^M)$ will only attend school when it is perceived as a feminine or neutral domain. The fact that school is perceived as masculine does not impact the net return to education for boys, so boys will (still) attend school whenever $\beta_i \geq \beta^M = \beta^*$. Hence in the M-equilibrium, the proportion of students who are female is:

$$P^M_{girls} = \frac{1 - F_\gamma(\gamma^M)}{2 - F_\beta(\beta^*) - F_\gamma(\gamma^M)}. \tag{B3}$$

The M-equilibrium exists if and only if $P^M_{girls} \leq \lambda$. Moreover, a sufficiently large ϕ (i.e. if girls face a sufficiently high psychic cost when entering a masculine domain) will guarantee that the M-equilibrium exists (holding λ fixed). At the same time, a higher λ (i.e. an increased inclination to perceive a school with mostly male students as a masculine domain) will increase the range of parameters that can support the M-equilibrium. Thus, if grammatical gender strengthens the inclination to perceive domains as either masculine or feminine through either theoretical mechanism, it will increase the scope for a more gender-segregated educational equilibrium.

The proportion of children who attend school is $1 - \left[F_\beta(\beta^*) + F_\gamma(\gamma^M)\right]/2$ in the M-equilibrium . Since $F_\gamma(\gamma^M) > F_\gamma(\gamma^*)$, fewer children attend school in the M-equilibrium than in the N-equilibrium. The F-equilibrium – in which school is a feminine domain – is defined symmetrically.

It is apparent that $P^M_{girls} \leq P^N_{girls} \leq P^F_{girls}$. Hence, multiple equilibria are possible whenever $P^M_{girls} \leq \lambda \leq P^N_{girls}$ or $P^N_{girls} \leq 1-\lambda \leq P^F_{girls}$. To see that at least one equilibrium always exists, first note that the N-equilibrium always exists if $\lambda \leq P^N_{girls} \leq 1-\lambda$. If the N-equilibrium does not exist because $\lambda > P^N_{girls}$, then λ must also be greater than P^M_{girls} – so the M-equilibrium exists. Similarly, if the N-equilibrium does not exist because $P^N_{girls} > 1-\lambda$, then $1-\lambda$ must also be less than P^F_{girls} – so the F-equilibrium exists.

Thus, 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. 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.

B.2 Labor Force Participation and the Division of Household Tasks

Next, we consider the decision problem facing two parents (one male and one female) 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 $\beta_i \sim F_\beta$.[33] γ and β both have continuous support between 0 and some finite maxima, β^{max} and γ^{max}.

A (heterosexual) household maximizes consumption:

$$C = w_{mom}L_{mom} + w_{dad}L_{dad} - w_{sitter}H_{sitter} \tag{B4}$$

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_{sitter} represents the market wage paid to babysitters. We assume that babysitters are female, and that they are young women who would not be included in the adult labor force if they were not employed sitters (for example, au pairs, older sisters).[34] Both mom and dad have one unit of time which they allocate to either work

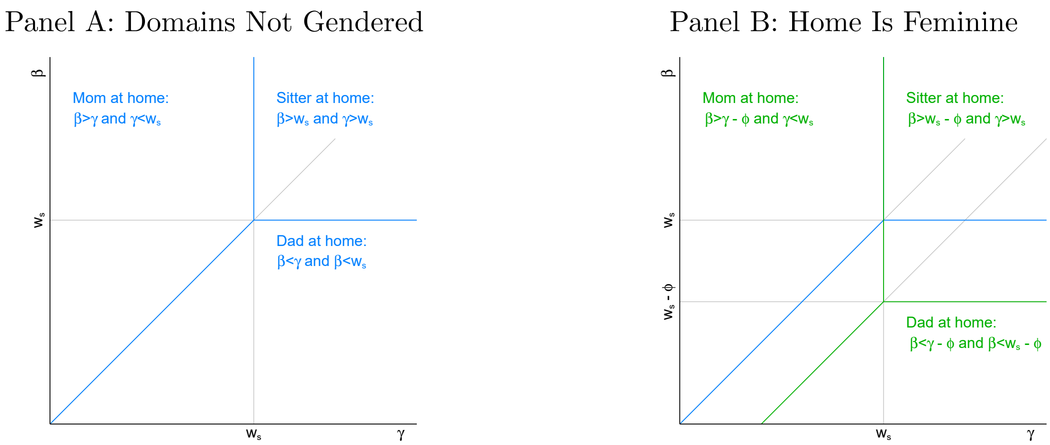
[33]For obvious reasons, using the subscripts m and f to distinguish between male and female adults who are also mothers and fathers might lead to confusion.

[34]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

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: $H_{mom} + H_{dad} + H_{sitter} = 1$.

First, consider the case where there are no gendered domains. A household will hire a sitter to take care of the children whenever both the mother and the father are both able to earn more than the sitter's wage – i.e. when $\beta_i \geq w_s$ and $\gamma_i \geq w_s$. When $\gamma_i < w_s$ and $\gamma_i \leq \beta_i$, the mother stays home while the father works. When $\beta_i < w_s$ and $\gamma_i > \beta_i$, the father stays home while the mother works. Panel A of Figure B1 illustrates this partition of the space of possible parental ability levels.

Figure B1: 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:

$$P^*_{mom} = \int_{\beta=0}^{\beta=w_s} \int_{\gamma=0}^{\gamma=\beta} f_{\beta,\gamma}(\beta, \gamma) + \int_{\beta=w_s}^{\beta=\beta^{max}} \int_{\gamma=0}^{\gamma=w_s} f_{\beta,\gamma}(\beta, \gamma). \tag{B5}$$

In other words, P^*_{mom} is the integral of $f_{\beta,\gamma}(\beta, \gamma)$ over the "mom at home" region in Figure B1. P^*_{dad} and P^*_{sitter} can be defined analogously:

$$P^*_{dad} = \int_{\beta=0}^{\beta=\gamma} \int_{\gamma=0}^{\gamma=w_s} f_{\beta,\gamma}(\beta, \gamma) + \int_{\beta=0}^{\beta=w_s} \int_{\gamma=w_s}^{\gamma=\gamma^{max}} f_{\beta,\gamma}(\beta, \gamma) \tag{B6}$$

and

$$P^*_{sitter} = \int_{\beta=w_s}^{\beta=\beta^{max}} \int_{\gamma=w_s}^{\gamma=\gamma^{max}} f_{\beta,\gamma}(\beta, \gamma). \tag{B7}$$

For any $f_{\beta,\gamma}(\beta, \gamma)$, $P^*_{mom} + P^*_{dad} + P^*_{sitter} = 1$ since households must either have mom, dad, or a sitter at home with the children. Since all households have exactly one person at home, the proportion 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.

homes where a woman takes care of the children is $P^*_{mom} + P^*_{sitter}$. The proportion of women in the (out-of-the-home) workforce is:

$$\frac{P^*_{dad} + P^*_{sitter}}{1 + P^*_{sitter}} \tag{B8}$$

since every household sends at least one adult into the workforce, and those with sitters send two. $(P^*_{mom}, P^*_{dad}, P^*_{sitter})$ 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.

For each candidate equilibrium HW, we define P^{HW}_{mom}, P^{HW}_{dad}, and P^{HW}_{sitter} as the proportion of households where (respectively) the mother, the father, or a sitter stays home with the children. $P^{HW}_{mom} + P^{HW}_{dad} + P^{HW}_{sitter} = 1$. We then define $Q^{HW}_{home} = P^{HW}_{mom} + P^{HW}_{sitter}$ as the proportion of households where the person (at home) taking care of the children is female. In equilibrium, home is perceived as a masculine domain if $Q^{HW}_{home} < \lambda$, and home is perceived as a feminine domain if $Q^{HW}_{home} > 1 - \lambda$. We define

$$Q^{HW}_{work} = \frac{P^{HW}_{dad} + P^{HW}_{sitter}}{1 + P^{HW}_{sitter}}$$

as the proportion of the out-of-home workforce that is female (note that households that hire a sitter send both a man and a woman into the out-of-home labor force, while other households send either a man or a woman). Work is perceived as a masculine domain whenever $Q^{HW}_{work} < \lambda$; it is perceived as a deminine domain whenever $Q^{HW}_{work} > 1 - \lambda$.

We begin by documenting a useful algebraic inequality that we will invoke repeatedly in our subsequent exposition. Trivial as it is, we label this inequality to avoid unnecessary repetition in our proofs and discussion.

Inequality 1. *For any $a \in (0, 1)$ and $b > 0$, $a < (a + b)/(1 + b)$.*

A direct consequence of Inequality 1 is that $P^{HW}_{dad} < Q^{HW}_{work}$: the proportion of households where the father is responsible for childcare is lower than the proportion of women in the out-of-home workforce. This follows from the definition of Q^{HW}_{work} above.

In our first set of results, we show that three of the candidate equilibria – NM, MN, and MF – cannot exist. As the lemmas below demonstrate, this does not depend on the strength of the inclination to partition the world into masculine and feminine domains (λ) or the magnitude of the cost of entering a domain dominated by the opposite sex (ϕ). Importantly, it indicates that whenever the tendency to partition enables the existence of a masculine or male-dominated domain, there must also be a female-dominated domain (in our inherently symmetric model, where the tendency to partition constrains both men and women).

Lemma 1. *If home is masculine, then work must be feminine.*

Proof. Home is perceived as masculine if and only if $Q^{HW}_{home} < \lambda$. This places bounds on the feasible range of values of Q^{HW}_{work}:

$$\begin{aligned} Q^{HW}_{home} < \lambda &\Leftrightarrow P^{HW}_{mom} + P^{HW}_{sitter} < \lambda \\ &\Leftrightarrow 1 - P^{HW}_{dad} < \lambda \\ &\Leftrightarrow P^{HW}_{dad} > 1 - \lambda \\ &\Rightarrow Q^{HW}_{work} > 1 - \lambda \end{aligned}$$

with the last step following immediately from Inequality 1 and the definition of Q^{HW}_{work}. □

Lemma 2. *If work is masculine, then home must be feminine.*

Proof. Work is masculine if and only if $Q^{HW}_{work} < \lambda$. Inequality 1 tells us that $Q^{HW}_{work} > P^{HW}_{dad}$, so $P^{HW}_{dad} < \lambda$ or, equivalently, $1 - P^{HW}_{dad} = P^{HW}_{mom} + P^{HW}_{sitter} = Q^{HW}_{home} > 1 - \lambda$. □

Lemma 1 and Lemma 2 prove that the MN, MM, and NM equilibria cannot exist. In our next piece of analysis, we will demonstrate that an equilibrium always exists and the multiple equilibria are possible. Before doing so, we introduce some additional notation. Partition the $\gamma \times \beta$ space into 13 regions a, b, c, d, e, f, g, h, i, j, k, l, and m, as in Figure B2. For capital letters $Z = A, B, C, \ldots M$, define Z as the integral of $f_{\gamma,\beta}(\gamma,\beta)$ over the region z. So,

$$A = \int_{\beta=2\phi}^{\beta=w_s+\phi} \int_{\gamma=0}^{\gamma=\beta+2\phi} f_{\beta,\gamma}(\beta,\gamma) + \int_{\beta=w_s+\phi}^{\beta=\beta^{max}} \int_{\gamma=0}^{\gamma=w_s-\phi} f_{\beta,\gamma}(\beta,\gamma) \tag{B9}$$

is the integral of $f_{\gamma,\beta}(\gamma,\beta)$ over the region a, and

$$B = \int_{\beta=w_s+\phi}^{\beta=\beta^{max}} \int_{\gamma=w_s-\phi}^{\gamma=w_s} f_{\beta,\gamma}(\beta,\gamma) \tag{B10}$$

is the integral of $f_{\gamma,\beta}(\gamma,\beta)$ over b.

Note that for any $f_{\gamma,\beta}(\gamma,\beta)$,

$$A + B + C + D + E + F + G + H + I + J + K + L + M = 1. \tag{B11}$$

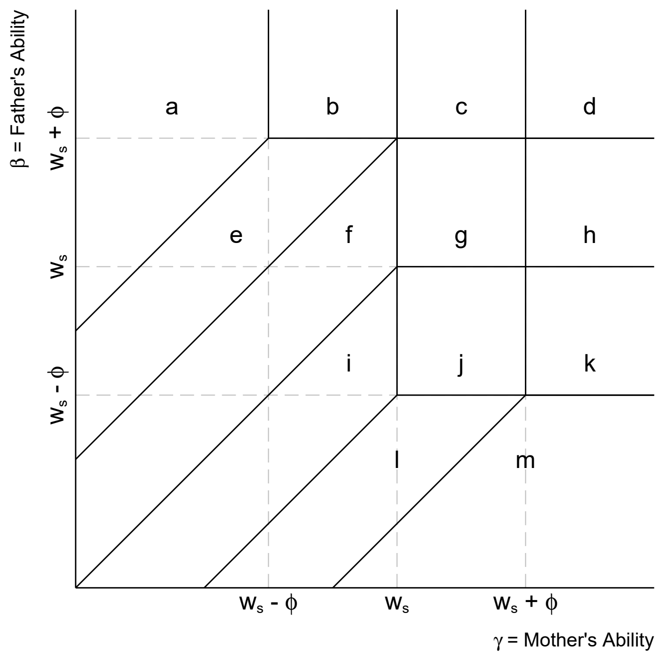
Moreover, for every candidate equilibrium HW, the proportion of households where a woman does the childcare, Q^{HW}_{home}, and the proportion of the workforce that is female, Q^{HW}_{work}, can be expressed in terms of A, B, C, etc.

Before proceeding to our existence proof, we comment on two particularly salient candidate equilibria: the NN equilibrium, in which neither work nor home is a gendered domain, and the FM equilibrium, in which work is a masculine domain and home is a feminine 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 B1): both parents work whenever $\gamma_i > w_s$ and $\beta_i > w_s$, 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 $P^{NN}_{sitter} = P^{*}_{sitter}$. However, when domains can be gendered, this is only an equilibrium when

$$\lambda < Q^{NN}_{home} < 1 - \lambda \tag{B12}$$

Figure B2: Labor Force Participation when Domains Are Not Gendered



and

$$\lambda < Q^{NN}_{work} < 1 - \lambda. \tag{B13}$$

Thus, 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, i.e. as the inclination to partition the world into masculine vs. feminine becomes more pronounced, limiting the scope for non-gendered domains. However, the existence of the NN equilibrium does not depend on the magnitude of the magnitude of the cost of entering a domain that does not align with one's own gender identity.

Next, we consider the FM equilibrium: home is a feminine domain and work is a masculine one. This equilibrium exists whenever

$$Q^{FM}_{home} = 1 - M > 1 - \lambda \tag{B14}$$

and

$$Q^{FM}_{work} = (D + H + K + M) \,/\, (1 + D + H + K) < \lambda. \tag{B15}$$

Clearly, as λ gets closer to one half, the range of parameter combinations that can support the existence of the FM equilibrium expands. However, is also the case that a larger psychic cost of entering a domain contrary to one's gender identity (ϕ) will increase the likelihood than the FM equilibrium exists (holding all other parameters equal). To see this, note that $1-M$ is (weakly) decreasing in ϕ, as is $D+H+K+M$. $1 + D + H + K$ can be increasing or decreasing in ϕ, but any decrease is no larger than the decrease in $D + H + K + M$ – so Q^{FM}_{work} is also decreasing in ϕ. Thus, an increased inclination to view the world in terms of masculine and feminine domains, whether manifest through an increase in ϕ or a λ closer to one half, increases the likelihood of an FM equilibrium characterized by traditional gender roles.[35]

We now conclude by demonstrating the existence of an equilibrium.

Proposition 1. *An equilbrium exists.*

Proof of Proposition 1. We structure the proof as follows. First, we show that when $A+B+C+D<\lambda$, the MF equilibrium always exists. Second, we show that whenever $A+B+C+D+E+F+G+H>1-\lambda$, at least one equilibrium exists in which home is a feminine domain. Finally, we consider the intermediate case where $A+B+C+D\geq\lambda$ but $A+B+C+D+E+F+G+H\leq 1-\lambda$; we show that either the NF or the NN equilibrium will always exist in this intermediate case.

1. Case 1: $A+B+C+D<\lambda$.
   First, note that $Q^{MF}_{home}=A+B+C+D$, so $Q^{MF}_{home}<\lambda$ (i.e. home is masculine). Next, note that

   $$\begin{aligned} Q^{MF}_{home}<\lambda &\Rightarrow 1-Q^{MF}_{home}>1-\lambda \\ &\Rightarrow P^{MF}_{dad}>1-\lambda \\ &\Rightarrow Q^{MF}_{work}>1-\lambda \end{aligned}$$

   by Inequality 1 and the definition of Q^{MF}_{work}. So, $A+B+C+D<\lambda$ implies that $Q^{MF}_{home}<\lambda$ and $Q^{MF}_{work}>1-\lambda$; hence, $A+B+C+D<\lambda$ implies that the MF equilibrium exists.

2. Case 2: $A+B+C+D+E+F+G+H>1-\lambda$.
   First, note that $A+B+C+D+E+F+G+H>1-\lambda$ implies that $Q^{FF}_{home}>1-\lambda$, $Q^{FN}_{home}>1-\lambda$, and $Q^{FM}_{home}>1-\lambda$ (this follows directly from the definitions of Q^{FF}_{home}, Q^{FN}_{home}, and Q^{FM}_{home} which are listed above). Next, note that if $Q^{FM}_{work}<\lambda$, then the FM equilibrium exists. If this condition doesn't hold,

   $$\begin{aligned} Q^{FM}_{work}>\lambda &\Leftrightarrow \frac{M+(D+H+K)}{1+(D+H+K)}>\lambda \\ &\Rightarrow \frac{M+(D+H+K)+(C+G+J)}{1+(D+H+K)+(C+G+J)}>\lambda \\ &\Rightarrow \frac{L+M+(D+H+K)+(C+G+J)}{1+(D+H+K)+(C+G+J)}>\lambda \end{aligned}$$

   and this final fraction is equal to Q^{FN}_{work}. If Q^{FN}_{work} is also less than $1-\lambda$, then the FN equilibrium exists. Finally, if $Q^{FN}_{work}>\lambda$ (so the FN equilibrium does not exist), we know that the FF equilibrium must exist because:

   $$\begin{aligned} \frac{C+D+G+H+J+K+L+M}{1+C+D+G+H+J+K}>\lambda &\Rightarrow \frac{C+D+G+H+I+J+K+L+M}{1+C+D+G+H+J+K}>\lambda \\ &\Rightarrow \frac{C+D+G+H+I+J+K+L+M}{1+C+D+G+H}>\lambda \end{aligned}$$

   which guarantees that the FF equilibrium exists because the last fraction is equal to Q^{FF}_{work}. So,

[35] It is important to note, however, that the likelihood that an MF equilibrium, where home is masculine and work is feminine, will also be generally higher when ϕ is larger or a λ closer to one half.

either the FF, the FN, or the FM equilibrium must exist whenever $A+B+C+D+E+F+G+H > 1-\lambda$.

Case 3: $A+B+C+D \geq \lambda$ and $A+B+C+D+E+F+G+H \leq 1-\lambda$.
First, note that $A+B+C+D \geq \lambda$ guarantees that $Q^{NF}_{home} \geq \lambda$ and $Q^{NN}_{home} \geq \lambda$ (and the equalities are strict when all regions a, b, c, etc. are non-empty). So, the requirement that home is a neutral domain in the NF and NN equilibria is satisfied. If it is also the case that

$$\frac{C+D+G+H+I+J+K+L+M}{1+C+D+G+H} \leq 1-\lambda$$

then $Q^{NN}_{work} \leq 1-\lambda$ (by the definition of Q^{NN}_{work}), so the NN equilibrium exists. However, if

$$\frac{C+D+G+H+I+J+K+L+M}{1+C+D+G+H} > 1-\lambda$$

then $Q^{NF}_{work} > 1-\lambda$ because

$$\begin{aligned}
&\frac{C+D+G+H+I+J+K+L+M}{1+C+D+G+H} > 1-\lambda \\
&\quad\Rightarrow \frac{(C+D+G+H+I+J+K+L+M)+F}{1+C+D+G+H} > 1-\lambda \\
&\quad\Rightarrow \frac{(C+D+G+H+I+J+K+L+M)+F}{1+C+D} > 1-\lambda.
\end{aligned}$$

So, when $A+B+C+D \geq \lambda$ and $A+B+C+D+E+F+G+H \leq 1-\lambda$, then either the NN equilibrium or the NF equilibrium must exist. □