

Gendered Language and Women’s Agency

Raine Spearman¹ Naveen Sunder¹²

This version: September 4, 2026

Abstract

Many languages have sex-based or gendered grammar, that is, they require speakers to mark male/female distinctions as part of regular everyday conversations. For example, in spanish book is masculine (*el libro rojo*), while table is feminine (*la mesa roja*). We ask whether women who speak such gendered languages report different gender attitudes and household outcomes. Using data on more than 1.5 million women residing in 53 countries and speaking 151 languages, we find that women speaking gendered languages report more intimate partner violence (IPV) (1.2 percentage points) and lower participation in household decisions (4–6 percentage points). This pattern is consistent with linguistic theory: repeated gender marking in everyday conversations may make gender categories more salient, and hence shape real world outcomes. We then examine whether the association operates primarily through long-run cultural transmission (native language) or through more immediate cognitive salience (interview language). We find evidence consistent with both channels, with somewhat stronger support for the cultural-transmission channel. We also show that the pattern remains after controlling for different cultural and institutional factors.

Keywords: Gender norms; Intimate partner violence; Sex-based grammar; Gendered language; Women’s empowerment; Low- and middle-income countries

JEL Codes: J12, J16, Z13, D83, O15

¹Department of Economics, Bentley University, Waltham, MA, USA.

²Corresponding author: nsunder@bentley.edu. We thank Michael Quinn, the participants at the Texas Tech University Undergraduate Research Conference, the Bentley Honors Conference, and the 95th Annual Meeting of the Southern Economic Association in Tampa, FL, for helpful suggestions. All remaining errors are our own.

1 Introduction

Intimate partner violence and limited household agency are two of the most persistent barriers to women’s development in low- and middle-income countries (LMICs). Globally, roughly one in three women experience physical or sexual violence from an intimate partner; across the developing world, many women report limited say in decisions about their own healthcare, finances, or family mobility (Shah and Barski, 2026; Jayachandran and Voena, 2026). The United Nations’ Sustainable Development Goal 5 makes gender equality a central development objective: it targets the elimination of violence against women (target 5.2) and women’s full participation in decision-making (target 5.5), yet both goals remain substantially unmet. Research has linked these inequalities to income, women’s control over resources, employment opportunities, household bargaining power, legal institutions, education, and community norms (Cools and Kotsadam, 2017; Hidrobo and Fernald, 2013; Heath, 2014; Duflo, 2012; Jayachandran, 2015; Shah and Barski, 2026; Jayachandran and Voena, 2026). This paper asks whether an enduring part of the social environment, language, also matters in shaping these outcomes. We focus on the grammatical structure of language, more specifically on how it codes gender categories - more specifically how different languages mark male/female distinctions when speakers use nouns, pronouns, adjectives, verbs, or agreement patterns. In Hindi, for example, descriptions of people, actions, and attributes are often gendered; in Amharic, grammatical forms can mark the gender of speakers. These are languages that we would consider as gendered (or ones that have a sex-based grammatical structure). An important feature of such languages is that they make gender a part of day-to-day speech. These languages, by making gender salient in everyday conversations, possibly shape real-world outcomes. Specifically, we address the following question: Is speaking a language with sex-based grammatical gender ³ associated with higher rates of women’s IPV experience and lower household decision-making participation?

We test this using 126 rounds of the Demographic and Health Surveys (DHS), nationally representative household surveys designed for cross-country comparison (The DHS Program, 2026a). Our analysis includes over 1.5 million women aged 15–49 across 53 countries and 151 unique languages. The countries represented in the analysis account for 52.3 percent of the global LMIC population (World Bank, 2026a,b). We merge respondents’ primary language to independent typological classifications of whether that language has sex-based grammatical gender, drawing primarily on the World Atlas of Language Structures (WALS). The baseline specification includes country, broad-religion, interview-year, and survey-cluster fixed effects, along with individual and household controls. Thus, the coefficient of interest – the association between sex-based grammatical gender and each outcome – is identified from variation in language spo-

³In this manuscript we use the terms “sex-based grammatical gender”, “sex-based gendered language” and “gendered language” interchangeably.

ken among women observed within the same survey clusters and country-year-religion cells. This is a meaningful source of variation: 74 percent of the final sample resides in a district or region where more than two languages are spoken. We show that the results are robust to using country or province fixed effects instead.

We have two main findings. First, women whose primary language is gendered report 1.2 percentage point higher rates of any IPV experience, or 15 percent of the sample mean, with increases across every type of IPV (physical, sexual, and emotional) for which data is collected. Second, women speaking gendered languages are less likely to participate in household decisions – participation is 6.6 percentage points lower for healthcare decisions, 4.6 percentage points lower for daily purchases, 5.3 percentage points lower for major purchases, and 6.5 percentage points lower for family visits, or about 7.4–10.0 percent of the relevant outcome means. We interpret these estimates with caution – they are not the *causal* relationship between gendered language use on IPV, but rather provide robust evidence of an association that is consistent with the hypothesis that socio-economic outcomes are shaped by the social environment in which they occur. The policy interpretation is that language is not a direct policy lever in the way that social transfers, legal protection, employment access, or schooling are, but it matters significantly (and needs to be accounted for) because it is part of the cultural setting in which those policies are implemented, and hence may shape their effects. This interpretation also speaks to debates over gender-fair and gender-inclusive language. Such reforms generally aim to reduce masculine-default wording, make women more visible, or avoid stereotype-reinforcing formulations, and prior work finds effects on gender representation, male salience, and stereotypes (Bigler and Leaper, 2015; Sczesny et al., 2016; Tavits and Pérez, 2019).

To understand the mechanisms driving our findings, we examine two possible channels. The first is *cultural transmission*: gendered grammar may be part of the inherited social environment through which people learn gender roles (Jayachandran, 2015), much as the legacy of plough-based agriculture persists in gender norms today (Alesina et al., 2013). The second is *cognitive salience*: the language used in a particular setting may make male and female categories more noticeable in the moment and hence change responses to questions about gendered norms and roles (Chen, 2013; Boroditsky et al., 2003). The former is a more persistent form of transmission, while the latter is more context-dependent (or immediate), and we examine these channels separately. First, we show that the pattern we observe is unique to gender-related attitudes and is absent for non-gender-specific beliefs and for generic violence-related attitudes. This suggests a strong *cultural transmission* channel. Second, we compare the associations with native language and interview language. If the immediate language of the interview were the main source of cognitive salience, interview language should matter more than native language. We find that both matter, with the native-language association somewhat stronger.

The evidence therefore points more strongly to the *cultural transmission* channel, while leaving room for the *cognitive salience* channel.

The paper makes three contributions. First, it contributes to the economics literature on IPV and women’s agency. This literature shows that IPV and household decision-making are shaped by income, women’s resources, outside options, laws, institutions, bargaining power, and gender norms (Cools and Kotsadam, 2017; Hidrobo and Fernald, 2013; Heath, 2014; Shah and Barski, 2026; Jayachandran and Voena, 2026). A related literature in economics studies culture as transmitted beliefs and values, and shows that gender norms can persist through families, communities, inherited institutions, and social learning (Guiso et al., 2006; Bisin and Verdier, 2001; Giuliano, 2020; Alesina et al., 2013; Jayachandran, 2015, 2021). Historical family systems and gender institutions have also been linked directly to violence against women and IPV (Alesina et al., 2021; Tur-Prats, 2019). Our contribution is to add language structure to this discussion. We demonstrate that language may be one durable channel through which gender hierarchy remains salient in the settings where household decisions and marital conflict occur.

Second, the paper contributes to research on identity and salience. In economics, identity can shape preferences and behavior (Akerlof and Kranton, 2000; Shayo, 2020), and making an identity salient can change choices and performance even when material incentives do not change (Hoff and Pandey, 2006; Benjamin et al., 2010; Cohn et al., 2014). Psychology and psycholinguistics make a related point: gendered language can make gender categories more visible, strengthen binary categorisation, and shape stereotypes and social representation (Bigler and Leaper, 2015; Sczesny et al., 2016; Gygax et al., 2019). Our contribution is to study a form of salience that is not created by an experiment or a one-time prompt. Sex-based gendered language is learned early, used repeatedly, and appears in everyday interactions within families, schools, markets, and public life. This also complements studies of gender-fair language interventions, which manipulate pronouns or wording in a specific setting and find effects on attitudes or performance in some domains but weaker evidence for immediate economic behavior (Pérez and Tavits, 2019; Kricheli-Katz and Regev, 2021; Balafoutas et al., 2025). This allows us to ask whether a repeated, real-world source of gender salience is associated with women’s outcomes in the household.

Third, the paper contributes to the economics of gendered language. Studies have shown that grammatical gender is related to gender gaps in labour-market, education, and attitudinal outcomes. Mavisakalyan (2015) links gender-intensive languages to female labour-force participation and discriminatory gender attitudes. Hicks et al. (2015) show that immigrants whose mother tongues place greater grammatical emphasis on gender allocate household tasks more strongly along sex lines. Gay et al. (2018) study married female immigrants in the United

States and show that sex-based grammatical rules are associated with lower labour supply, partly through broader cultural factors related to household bargaining and ethnic enclaves. [Jakiela and Ozier \(2020\)](#) expand the measurement of grammatical gender across thousands of languages and relate it to women’s education, labour-force participation, and gender norms. The study most directly related to ours is [Davis et al. \(2024\)](#), who link sex-based grammatical gender to country-level IPV incidence and individual beliefs about the justifiability of IPV. We build on this evidence in three ways. First, we move beyond labour-market and education outcomes to safety and agency within marriage. Second, we use DHS microdata on reported IPV experience and household decision-making, rather than country-level IPV incidence or broad attitudinal outcomes. Third, we use nationally representative household surveys covering countries that account for more than 50% of the global LMIC population.

Section 2 reviews related literature. Sections 3 and 4 describe the data and empirical strategy. Section 5 reports the results and robustness checks. Section 6 examines mechanisms. Section 7 concludes.

2 Related Literature

Language, Cognition, and Gender Norms

The paper’s theoretical premise draws on a weak version of linguistic relativity: language does not determine thought, but grammatical structure can influence attention, categorisation, and salience ([Whorf, 1956](#); [Chen, 2013](#)). Grammatical gender is a natural setting for this hypothesis because many languages require speakers to classify nouns, people, and roles through gendered morphology. The evidence should be read carefully. [Samuel et al. \(2019\)](#) review empirical studies on grammatical gender and thought and conclude that effects are task- and context-dependent, with alternative explanations often plausible. This caveat is important for the present study: the claim is not that sex-based grammar mechanically produces gender hierarchy, but that repeated grammatical marking may make gender categories cognitively salient under some conditions. Recent economic evidence on language and behaviour reinforces this caution: [Keller et al. \(2024\)](#) find no consistent relationship between endogenous future-time reference and patience, suggesting that linguistic effects depend on the domain, exposure margin, and behavioural outcome.

Experimental and psychological evidence supports this more cautious interpretation. [Boroditsky et al. \(2003\)](#) show that speakers of gendered languages more readily attribute gendered characteristics to objects, while [Cubelli et al. \(2011\)](#) demonstrate that grammatical gender affects object categorisation among Italian and Spanish speakers. [Sato et al. \(2020\)](#) show that

grammatical gender can enter early perceptual and attentional processing in gender-associated categorisation tasks. The broader gendered-language and gender-fair-language literature is especially relevant because it concerns people, roles, stereotypes, and social categories. [Bigler and Leaper \(2015\)](#) argue that gendered language can increase the salience of gender, reinforce binary categorisation, and encourage stereotypic views. Gender-fair or gender-inclusive language is usually discussed through two main strategies: neutralization, which replaces masculine-default forms with gender-unmarked forms, and feminization or pair forms, which explicitly make women visible ([Sczesny et al., 2016](#); [Gabriel et al., 2018](#)). Evidence from occupational and recruitment contexts shows that gendered wording can shape perceived belonging and the mental availability of women in male-typed roles ([Gaucher et al., 2011](#); [Vervecken et al., 2013](#); [Lindqvist et al., 2019](#)). At the same time, these effects vary by language, context, and outcome, and reviews caution against treating grammatical gender as a single homogeneous exposure ([Gygax et al., 2019](#); [Samuel et al., 2019](#); [Lange and von Stockhausen, 2025](#)).

The strongest evidence for social attitudes comes from experimental settings. [Pérez and Tavits \(2019\)](#) randomly assign bilingual Estonian–Russian speakers to respond in a genderless or gendered language and find that the genderless-language condition produces more egalitarian gender attitudes, with no comparable effect for attitudes without obvious gender referents. [Tavits and Pérez \(2019\)](#) show that gender-neutral pronouns in Swedish reduce male salience and shift attitudes toward women and LGBT individuals in public life. [Kricheli-Katz and Regev \(2021\)](#) move closer to behaviour, showing that addressing women in feminine rather than masculine grammatical forms in Hebrew reduces the gender gap in math performance and affects reported effort and stereotypes. Together, this literature suggests a plausible cognitive foundation for our channel claim: gendered language can make gender categories socially meaningful, and the DHS IPV-justification results are informative precisely because they concern gender-hierarchical beliefs rather than general violence attitudes.

IPV, Household Agency, and Norms

The two primary outcomes in this paper—reported IPV experience and household decision-making—are central development outcomes. IPV is a direct harm to women’s health, autonomy, and wellbeing, and it has intergenerational consequences for children and household resource allocation ([Shah and Barski, 2026](#)). Household decision-making captures a core dimension of agency in [Kabeer \(1999\)](#)’s empowerment framework: the ability to participate in choices over health, purchases, mobility, and family relationships. In collective household models, such participation is closely related to bargaining power, although DHS decision-making items remain proxies for agency rather than direct measures of Pareto weights ([Lundberg and Pollak, 1993](#); [Kishor and Subaiya, 2008](#); [Jayachandran and Voena, 2026](#)).

Development economics and public health research show that IPV is shaped not only by resources and law, but also by community norms. [Cools and Kotsadam \(2017\)](#), using DHS data from 30 Sub-Saharan African countries, show that resources are associated with lower acceptance of wife beating but do not always protect women from abuse, and that employed women face greater risk in communities with higher acceptance of wife beating. Cross-national DHS evidence similarly shows that domestic-violence acceptance is widespread across LMICs and varies systematically with national conditions ([Sardinha and Nájera Catalán, 2018](#)). More recent DHS studies find that wife-beating justification is associated with reported IPV experience in Uganda and in couple-level data from Malawi, Zambia, and Zimbabwe ([Kadengye et al., 2023](#); [Bengesai and Khan, 2023](#)). Systematic reviews of IPV in LMICs identify acceptance of gendered violence and patriarchal gender norms as important risk factors ([Ghoshal et al., 2022](#); [Gunaratne et al., 2023](#)). [Shah and Barski \(2026\)](#), in a comprehensive review of the economics of IPV in LMICs, likewise emphasize economic hardship and societal norms as central determinants.

This literature motivates our treatment of IPV justification. The DHS wife-beating questions do not ask about generic violence. They ask whether a husband is justified in hitting or beating his wife in scenarios involving marital authority, sexual availability, domestic compliance, and respect for husbands or in-laws. We therefore use IPV justification as a normative-channel measure: it captures the acceptability of gendered marital discipline, which is the kind of norm through which language might plausibly matter. At the same time, the measure must be interpreted cautiously. IPV justification is not evidence that victims’ own beliefs cause their abuse. It is better read as a marker of a broader normative environment in which IPV may be more tolerated, help-seeking may be inhibited, and women’s household bargaining position may be constrained.

A closely related strand of development research examines whether economic empowerment reduces IPV and improves decision-making. Cash-transfer studies document reductions in controlling behaviour and experienced violence among beneficiaries ([Hidrobo and Fernald, 2013](#)), while labour-market access is associated with improvements in household decision-making autonomy across domains ([Heath, 2014](#)). On household decision-making specifically, [Annan et al. \(2021\)](#) document substantial variation in DHS-measured participation across wealth quartiles in Sub-Saharan Africa, and [Jayachandran and Voena \(2026\)](#) show that participation in health-care, purchase, and family-visit decisions rises systematically with development across LMICs. These studies provide the benchmark for interpreting the magnitude of our estimates and clarify why household agency and reported IPV experience belong at the centre of the paper.

Gendered Language and Economic Behavior

The economics of culture documents that contemporary gender inequalities often reflect historical practices that persist long after their institutional origins have changed. [Alesina et al. \(2013\)](#) show that pre-industrial reliance on the plough predicts more regressive gender norms and lower female labour force participation today, including among descendants in settings with different institutions. [Jayachandran \(2015\)](#) surveys related persistence channels, including son preference, early marriage, restricted property rights, and occupational segregation, and [Jayachandran \(2021\)](#) shows that gender norms remain a binding constraint on women’s employment even at similar levels of development. Grammatical gender fits this tradition as a structural feature of inherited culture, but it is unusually pervasive because it is embedded in daily communication and acquired early in life.

A growing economics literature links grammatical gender to gendered economic behaviour and attitudes. [Mavisakalyan \(2015\)](#) finds that countries whose majority language is gender-intensive have lower female labour-force participation and that individual speakers of gender-intensive languages report more discriminatory gender attitudes. [Hicks et al. \(2015\)](#) provide closely related evidence in *Journal of Economic Behavior & Organization*, showing that gender marking in immigrants’ mother tongues predicts a more sex-specialised division of household labour. [Gay et al. \(2018\)](#) use an epidemiological approach among married female immigrants in the United States and find that sex-based grammatical rules are associated with lower female labour supply, with household bargaining and ethnic enclaves explaining part of the relationship. [Jakiela and Ozier \(2020\)](#) document grammatical gender across thousands of languages and relate it to female labour force participation, education, and gender norms, including in LMIC contexts. Other work links gendered language to educational attainment gaps, occupational belonging, and wage inequality ([Stout and Dasgupta, 2011](#); [Davis and Reynolds, 2018](#); [Van der Velde et al., 2015](#); [Cohen et al., 2023](#)). Recent work in *Journal of Economic Behavior & Organization* also links grammatical structure to family outcomes: [Lian et al. \(2026\)](#) find that pro-drop language use is associated with divorce behaviour and attitudes. This emerging evidence makes household bargaining, family stability, and marital norms natural margins on which to study the economic relevance of grammar.

A directly related paper is [Davis et al. \(2024\)](#), who link sex-based grammatical gender to country-level IPV incidence and individual-level beliefs about the justifiability of IPV, with evidence consistent with both cultural and cognitive channels. This means the contribution of the present paper is not the claim that grammatical gender predicts IPV beliefs. Rather, our contribution is to move from aggregate incidence and broad attitudinal evidence to individual-level reported IPV experience and household decision-making in a large, standardized DHS sample of married women in LMICs. By estimating these primary outcomes together with

IPV justification, we connect linguistic structure to women’s safety, household agency, and the gender-hierarchical normative environment that makes the pattern interpretable.

3 Data

3.1 DHS Data

Our individual-level outcome data come from the Demographic and Health Surveys (DHS), a large-scale, nationally representative, and standardised survey system covering low- and middle-income countries over more than three decades ([The DHS Program, 2026a](#)). The core questionnaire is administered using a stratified two-stage cluster sampling design. The current analysis sample covers 1,536,518 respondents across 126 distinct survey rounds, 53 countries, 151 unique languages, and 217 country-language pairs.⁴ We organize outcomes hierarchically. The primary outcomes are reported IPV experience, which captures women’s safety, and household decision-making participation, which captures a central dimension of household agency. We use IPV justification as a normative-channel measure.

Primary outcome: reported IPV experience. The DHS domestic violence module asks a randomly selected subsample of eligible women whether they have experienced specific forms of violence by a current or former husband or partner. We construct indicators for physical IPV, sexual IPV, emotional IPV, and any IPV. These outcomes are closer to realized harm than stated attitudes, but they remain reported measures: disclosure may be affected by fear, stigma, normalization of abuse, survey privacy, and local norms around speaking about marital violence ([The DHS Program, 2026c](#); [Shah and Barski, 2026](#)). We therefore interpret them as reported IPV experience rather than perfectly observed violence.

Primary outcome: household decision-making. We construct binary indicators for whether the woman has any say—individually or jointly with her spouse—in: (i) her own healthcare; (ii) daily household purchases; (iii) major household purchases; (iv) visiting family or relatives; and, where available, (v) staying with or visiting family. The current analysis data contain no nonmissing observations for the stay-family item, so the component results focus on the four available domains. These dimensions reflect key domains of intra-household bargaining power and female autonomy ([Kabeer, 1999](#); [Lundberg and Pollak, 1993](#)). In [Kabeer \(1999\)](#)’s three-dimensional empowerment framework, these outcomes capture *agency* — the capacity to make meaningful choices — which can remain constrained even when women’s *resources* and formal

⁴The DHS Program reports representative data from more than 300 surveys in over 90 countries ([The DHS Program, 2026b](#)). The domestic-violence module is optional and available only for some countries and survey rounds ([The DHS Program, 2026c](#)). We searched for the relevant IPV-justification and domestic-violence-module variables and, to our knowledge, use all available country-rounds with the required information that can be linked to the language classification.

legal *achievements* have improved. [Jayachandran and Voena \(2026\)](#) note that DHS decision-making questions ask whether a woman *makes* or jointly makes each type of decision, whereas a fuller measure of agency would capture whether she *influences* decisions — a distinction that matters because women who delegate decisions while retaining influence may score low on the DHS measure despite retaining some agency. We interpret our outcomes as proxies for decision-making roles rather than direct measures of the Pareto weight, consistent with their widespread use in the literature ([Kishor and Subaiya, 2008](#); [Annan et al., 2021](#)). Crucially, measurement error in the outcome variable attenuates OLS estimates toward zero — making our estimates conservative if the DHS questions understate true agency differences.

Normative-channel measure: IPV justification. Respondents are asked: “In your opinion, is a husband justified in hitting or beating his wife in the following situations?” Seven situations are presented: (1) she goes out without consent; (2) she neglects the house or children; (3) she argues with him; (4) she refuses sex; (5) she does not cook properly; (6) he suspects infidelity; (7) she shows disrespect toward his in-laws. Each response is coded as a binary indicator (1 = yes). We treat these variables as norm measures rather than as primary welfare outcomes. They are useful because the scenarios encode marital authority, sexual availability, domestic obligation, and respect for husbands or in-laws—the gender-hierarchical content through which grammatical gender may plausibly matter. Prior DHS work shows that acceptance of wife beating is widely used as a measure of domestic-violence norms and is empirically linked to IPV risk ([Sardinha and Nájera Catalán, 2018](#); [Cools and Kotsadam, 2017](#); [Bengesai and Khan, 2023](#); [Kadengye et al., 2023](#); [Shah and Barski, 2026](#)). We do not interpret these measures as evidence that victims’ own attitudes cause abuse; they proxy the broader normative environment around marital violence.

3.2 Language Data: WALS and Other Sources

Our primary source on linguistic structure is the World Atlas of Language Structures (WALS), a comprehensive cross-linguistic database compiled by the Max Planck Institute for Evolutionary Anthropology ([Corbett, 2013a,b,c](#)). WALS codes over 190 structural features for more than 2,600 languages. The main independent variable is *sex-based grammatical gender* (WALS feature 31A), which records whether a language’s gender system is organised around sex-based distinctions, such as masculine and feminine categories. Languages such as Spanish require grammatical gender marking across nouns and agreement patterns; English does not. We also use WALS feature 30A, which records the number of grammatical genders, and WALS feature 32A, which records whether gender assignment is primarily semantic, formal, or mixed, as auxiliary checks on the language measure.

WALS is the starting point because it is a widely used typological source based on descriptive

grammars and expert coding. However, DHS language labels do not always map directly onto a WALS entry. We therefore supplement WALS with source-backed linguistic evidence from Grambank, Glottolog, APiCS for pidgin and creole languages, and published grammars, grammar sketches, theses, and peer-reviewed linguistic descriptions (Michaelis et al., 2013; Skirgård et al., 2023; Haynie et al., 2023). Grambank is useful for languages outside direct WALS coverage because it provides systematically coded grammatical features for more than 2,400 languages. Glottolog helps identify language names, aliases, and classifications, while grammatical claims are based on WALS, Grambank, APiCS, or descriptive linguistic sources. This source hierarchy follows the broader economics and psychology literature that uses grammatical-gender measures to study gender attitudes and outcomes (Mavisakalyan, 2015; Jakiela and Ozier, 2020; Gygax et al., 2019).

3.3 Merging DHS with Language Data

To attach linguistic structure to individual DHS records, we construct a respondent-level primary language and then match that language to the source-backed grammatical-gender classification. This step is necessary because the DHS records language information differently across countries and survey rounds. Where available, we use native or mother-tongue language, since it best captures early linguistic socialisation. When native language is unavailable, we use interview language and, where necessary, questionnaire language or the dominant national language, guided by survey context and data quality. For India, where linguistic heterogeneity is especially important for the analysis, we use the DHS “Language of Respondent” variable.

The research team standardised DHS language labels across country-rounds before merging them to linguistic sources. This process corrected spelling variants, capitalisation differences, accented-character encodings, and country-specific label formats. It also separated valid language names from residual or aggregate labels, such as “other language,” numeric categories, and broad regional labels that could not be mapped to a single linguistic system. Each usable label was then matched to a canonical language identity using country context, known aliases, and linguistic identity sources.

The coding rule is conservative. A language is classified as sex-based when the source evidence indicates a grammatical gender system organised around masculine/feminine or otherwise sex-based distinctions. A language is classified as non-sex-based when the evidence indicates no grammatical gender or a non-sex-based noun-class system, such as many Bantu systems in which nouns are organised into multiple semantic or formal classes rather than a binary sex-based opposition. When direct evidence is unavailable or a DHS label bundles multiple languages, we either leave the observation outside the strict classification or use a separately flagged lower-confidence assignment only when the country context and closest-language evidence are strong.

The final crosswalk records the source type, citation or URL, confidence level, assignment scope, and review notes for each country-language label. It covers the 217 country-language pairs represented in the analysis and makes the grammatical-gender merge auditable. In the largest main-outcome specification, all 1,531,703 observations have a non-conflicting sex-based language assignment; 81.6 percent are covered by high-confidence, non-proxy assignments, and 91.2 percent are covered by high- or medium-confidence assignments. Restricting the sample to respondents whose primary language can be matched to the language data and who have nonmissing baseline controls yields the analysis sample. Appendix B describes the language-classification workflow in more detail.

India accounts for 321,816 observations, or roughly 21 percent of the main analysis sample, reflecting the scale of India’s National Family Health Survey (NFHS) and the rich language-of-respondent variable it provides.

In the Indian context, religious affiliation and primary language are correlated: Hindu respondents are disproportionately Indo-Aryan (gendered language) speakers, while Christian respondents in southern India are disproportionately Dravidian (non-gendered) speakers. The baseline specifications therefore absorb broad Hindu, Muslim, and Christian indicators, with other affiliations retained in the reference category. These indicators reduce confounding from religion-language correlations, though they may also absorb some meaningful social variation linked to language. We interpret the estimates as conditional descriptive associations rather than experimental effects.

3.4 Summary Statistics

Table 1 presents descriptive statistics in the reframed outcome order and restricts each row to the corresponding analysis sample used in Tables 2 or 3. The table first reports language exposure, then reported IPV experience, household decision-making, IPV-justification channel measures, and baseline controls. The DHS-language merge spans 53 countries, 151 unique language labels, and 217 country-language pairs, with large multilingual samples in countries such as India, Nigeria, Kenya, and Zambia. Approximately 46 percent of observations in the baseline IPV-justification sample speak a language with a sex-based grammatical gender system. The mean reported-IPV outcomes range from 2 percent for sexual IPV to 8 percent for any IPV in the domestic-violence-module sample. Participation rates in the available household-decision domains average roughly 61–68 percent. IPV-justification rates vary substantially across scenarios, from 15 percent for not cooking properly to roughly 30 percent for neglecting children.

The sample size varies across outcomes, reflecting differences in DHS questionnaire modules

across country-years. The first five IPV scenarios and most decision-making domains are asked in the large majority of survey rounds, while the infidelity and disrespect-to-in-laws scenarios are administered in only three countries in the current cleaned sample. We include interview-year fixed effects throughout to absorb calendar-time trends in attitudes and survey design changes. The core modules on IPV justification and household decision-making have been consistent across the DHS waves in our sample, with question wording standardised since the late 1990s (Kishor and Subaiya, 2008).

4 Empirical Strategy

Our empirical strategy compares married women observed in the same DHS survey clusters but speaking languages with different grammatical-gender structures. The exercise is descriptive: language is not randomly assigned, and the estimates should not be read as experimental effects. The goal is to ask whether, after absorbing local survey-cluster conditions and adjusting for observed respondent characteristics, sex-based grammatical gender is systematically associated with reported IPV experience, household decision-making, and IPV-justification norms.

We estimate:

$$Y_{isct} = \beta \cdot \text{GenLang}_{l(i)} + \mathbf{X}'_{isct}\boldsymbol{\gamma} + \alpha_c + \rho_r + \lambda_t + \delta_s + \varepsilon_{isct} \quad (1)$$

where Y_{isct} is a binary gender-related outcome for individual i in survey cluster s , country c , and interview year t ; $\text{GenLang}_{l(i)}$ is an indicator for whether respondent i 's primary language l has a sex-based gender system (WALS feature 31A); $\mathbf{X}_{isct}$ is a vector of individual and household controls; α_c are country fixed effects; ρ_r are broad religion fixed effects for Hindu, Muslim, and Christian affiliation, the largest reported affiliation groups in the sample; respondents outside these groups remain in the omitted or other category. λ_t are interview-year fixed effects; and δ_s are survey-cluster fixed effects. Standard errors are clustered at the survey-cluster level. The coefficient of interest, β , captures the association between sex-based grammatical gender and the outcome after comparing respondents within the same survey cluster and adjusting for country, broad religion, interview year, and observed respondent characteristics.

The controls are chosen to remove the most direct observable confounders. Survey-cluster fixed effects absorb local geography, local institutions, survey administration, and other conditions shared by women in the same DHS survey cluster. Interview-year fixed effects absorb common time patterns in survey design, attitudes, and reporting. Broad religion fixed effects absorb differences associated with Hindu, Muslim, and Christian affiliation. Individual controls include age and age squared, highest educational level, household wealth quintile, and an indicator for female household head. For example, if two married women are observed in the same survey cluster and differ in whether their primary language has sex-based grammatical gender, the

specification compares their outcomes after holding constant local cluster conditions, broad religious affiliation, survey year, age, education, wealth, and household-head status.

This empirical design controls for many confounders, but not all of them. It does not control for unobserved traits that vary across language communities within the same cluster, such as detailed ethnic histories, caste or kinship systems, partner characteristics, marriage-market selection, migration and residential sorting, local enforcement of laws, reporting norms, or other inherited cultural institutions correlated with language. The analysis is also restricted to married women observed in DHS modules that measure reported IPV experience, household decision-making, or IPV justification. If women in grammatically gendered-language communities enter or remain in marriage at different rates, sample selection may affect the estimates. For IPV justification, higher marriage rates in more patriarchal linguistic communities could overrepresent women who have internalised those norms. For household decision-making and reported IPV, the direction of selection is less clear. These limitations are why we interpret β as a conditional descriptive association rather than a causal effect.

However, we conduct a number of robustness checks to increase the strength of our findings. Reverse causality is unlikely: a respondent’s current IPV experience or household decision-making role cannot retroactively alter the grammatical structure of her language, which evolves over long periods. We also use another multi-country dataset (the WVS), to show that the association is stronger for gender-hierarchical attitudes than for general acceptance of violence or broad conservatism, and the future-tense placebo does not reproduce the main pattern. Additional robustness checks ask whether the results are sensitive to ethnicity fixed effects, province and country fixed effects, leaving out large countries one at a time, using only high-confidence language classifications, distinguishing native from interview language, adding historical institution controls, and instrumenting current grammatical gender with proto-language gender. These checks reduce concerns that the estimates are driven by arbitrary linguistic structure, source quality, one large country, or broad inherited institutions.

5 Results

5.1 IPV Experience and Household Agency

Table 2 reports the baseline estimates⁵. The table is organized around the paper’s two primary development outcomes: reported IPV experience, which captures women’s safety, and household decision-making, which captures agency within the family. Panel A uses the DHS domestic-violence module, administered to a randomly selected eligible respondent in partici-

⁵Figure 1 summarizes the coefficient pattern for the two primary outcome domains.

pating households. Women speaking a sex-based gendered language are more likely to report IPV experience across all four composites. Any IPV rises by 1.2 percentage points, physical IPV by 1.2 percentage points, sexual IPV by 0.3 percentage points, and emotional IPV by 0.4 percentage points. Relative to the corresponding sample means, these estimates are approximately 8-22 percent of the respective sample means. The magnitudes are modest in absolute terms, but they are meaningful for a low-base-rate harm outcome measured with substantial reporting constraints.

Panel B reports larger and more uniform associations for household decision-making. Speaking a sex-based gendered language is associated with a 6.2 percentage point reduction in the decision-making index. The component estimates are negative in every available domain: healthcare (-6.6 percentage points), daily purchases (-4.6 percentage points), major purchases (-5.3 percentage points), and visits to family or relatives (-6.5 percentage points). The stay-family item has no available observations in the current analysis data and is therefore not interpreted as a finding. These associations are consistent with lower household agency and bargaining power in domains where gendered expectations are especially salient.

[Table 3](#) reports compact IPV-related belief measures used as channel evidence and discussed in [Section 6](#). These outcomes are not primary harm or agency outcomes, but they are useful because they capture whether respondents accept wife beating in scenarios tied to marital authority, sexual availability, and domestic compliance.

The decision-making estimates are large relative to familiar benchmarks, so we interpret them with care. [Annan et al. \(2021\)](#) document 5–10 percentage point variation in DHS-measured household decision-making participation across wealth quartiles in 23 Sub-Saharan African countries. [Heath \(2014\)](#) finds that labour-market access in Bangladesh is associated with 6–12 percentage point improvements in household decision-making autonomy across domains, and [Duflo \(2012\)](#) surveys evidence that transfers targeted to women typically shift household decision-making participation by 3–8 percentage points. Our estimates fall in this range, but the comparison is not to a marginal policy intervention. Sex-based grammatical gender is a durable feature of the social environment. It is learned early, used repeatedly, and bundled with the gender norms, expectations, and community histories attached to language groups. A larger association is therefore plausible because the coefficient may capture cumulative exposure to a linguistic and normative environment, not a short-run change in one resource or program.

5.2 Heterogeneity

The main estimates also vary across respondent contexts in ways that help describe where the association is strongest. These splits are not mechanism tests on their own. They show

whether the language association is larger among groups with less exposure to schooling, formal institutions, and alternative gender norms.

This comparison is useful because education and residence are central margins in the development literature on gender norms and IPV. Schooling can do more than raise earnings. It can expose adolescents to peers, teachers, curricula, and public institutions that make inherited gender rules easier to question. Consistent with this view, a school-based intervention in India that used classroom discussions about gender equality shifted adolescents' attitudes in a more egalitarian direction and produced some changes in self-reported behaviour (Dhar et al., 2022). Residence may matter for a related reason. Urban areas often provide denser contact with labour markets, media, services, and non-kin networks; in rural India, for example, the introduction of cable television reduced reported acceptance of domestic violence and increased women's autonomy (Jensen and Oster, 2009). Larger multi-country studies also show that IPV acceptance and IPV exposure are often higher among less-educated and rural women (Tran et al., 2016; Gunarathne et al., 2023; Nabaggala et al., 2021). Our heterogeneity exercise therefore asks whether the grammatical-gender association is strongest in the same settings where prior work finds gender hierarchy to be more persistent.

Figure 2 reports education heterogeneity by estimating equation (1) separately for women with five or fewer years of schooling and those with more schooling. Decision-making coefficients are negative in both education groups but generally larger for less-educated respondents, while reported-IPV coefficients are concentrated among the lower-education group. IPV-justification heterogeneity is more mixed across components, consistent with the smaller full-sample channel estimates in Table 3. In light of the schooling literature, this pattern is informative because the language association attenuates among women who are more likely to have encountered institutional and peer contexts outside the family. The result does not show that education is the mechanism. It shows that the association is strongest where the social environments that transmit inherited norms may be less likely to be counterbalanced by formal schooling.

Figure A1 reports urban-rural heterogeneity. The decision-making coefficients are negative in both settings but generally larger in rural areas. Reported IPV experience is also concentrated in rural areas: physical and any-IPV coefficients are positive and significant among rural respondents but close to zero in urban samples. IPV-justification heterogeneity is less uniform, with some scenarios showing larger rural coefficients and others differing in sign across residence groups. This rural concentration is also consistent with the broader evidence that rural residence often coincides with lower institutional density, weaker access to outside information, and stronger local enforcement of family and marital norms. At the same time, the mixed IPV-justification components caution against a simple story in which urban residence uniformly weakens every gendered belief. Taken together, the education and residence splits

contextualize the main estimates rather than prove a separate mechanism: gendered grammar is most predictive where prior work would lead us to expect fewer countervailing exposures to alternative gender norms.

5.3 Robustness Checks

Same-sample robustness. [Table A1](#) addresses a sample-composition concern. Reported IPV is measured only for women selected into the DHS domestic-violence module, while decision-making and IPV-justification outcomes are observed in broader DHS samples. If the outcome hierarchy were driven mainly by differences across respondent pools, the pattern should weaken when outcomes are compared within the domestic-violence-module sample. The same-sample exercise therefore restricts the comparison to a more similar respondent pool where possible. Reported-IPV coefficients remain positive, and household-decision coefficients remain negative and somewhat larger than in the broader decision-making sample: the decision index falls by 7.8 percentage points, with declines of 8.7 percentage points for healthcare, 6.5 percentage points for major purchases, and 8.3 percentage points for family visits. The IPV-justification index is close to zero in this restricted sample, although selected component outcomes remain positive for arguing, refusing sex, and not cooking. This check therefore supports the primary harm and agency results while preserving the interpretation of the belief estimates as evidence on the surrounding gender-norm environment.

Unrelated-language placebo. A second concern is that the estimates could reflect broad differences between language communities rather than sex-based grammatical gender specifically. [Figure A2](#) replaces grammatical gender with a linguistic feature unrelated to gender: grammatical future tense. The placebo is conducted using the corresponding baseline controls and fixed effects where supported by the outcome sample. The placebo coefficients do not reproduce the decision-making and IPV-experience pattern from the main grammatical-gender estimates. This helps address the concern that any salient linguistic structure, or any language-community difference captured by WALs, would generate the same empirical pattern.

Ethnicity, geography, and source quality. Related languages may also share ethnic histories, regional institutions, or source-quality features independently of the social process of interest. To address this concern, we estimate sensitivity checks that absorb ethnicity fixed effects where available and compare specifications with province and country fixed effects. We also restrict the sample to higher-confidence language classifications. [Figure 3](#) gives a compact summary of these baseline, ethnicity-FE, province-FE, country-FE, and high-confidence checks for the main summary outcomes and the IPV-justification channel index. [Figure A3](#), [Figure A4](#), and [Figure A5](#) report the component-level fixed-effect sensitivity checks. [Figure A6](#) reports the

source-quality restriction. Across these checks, household decision-making remains consistently negative, and reported IPV experience remains positive in the principal specifications. The smaller IPV-justification coefficients are more mixed across components, which is consistent with their role as mechanism evidence. These results reduce the concern that the main hierarchy is an artifact of ethnicity coding, geographic fixed-effect choice, or lower-confidence language matches.

Historical-institution controls. A related concern is that sex-based grammatical gender may proxy for deeper historical gender institutions rather than capturing a distinct linguistic association. We therefore merge the DHS country-language analysis cells to ancestral society-level measures from D-PLACE and the Ethnographic Atlas (Kirby et al., 2016; Murdock, 1967; D-PLACE, 2026). Table 4 adds controls for ancestral plough use, descent systems, post-marital residence, bride price, and dowry.⁶ In the matched sample the coefficients agree in sign with the baseline estimates: reported IPV and IPV justification are positive and household decision-making is negative. The decision-making association remains strongly negative after adding the historical controls, including the extended specification with plough, descent, residence, bride price, and dowry. This mitigates the concern that the decision-making result is mechanically inherited from those historical institutions. The reported-IPV and IPV-justification estimates are more sensitive: both lose significance once descent is included, and both fall close to zero under the joint extended controls. That is informative rather than fatal. It suggests that language and older gender institutions are partly bundled in the historical record, while the decision-making result continues to survive this demanding comparison.

Instrumental variables. A further concern is that current sex-based grammatical gender could be bundled with other durable community traits. Table 5 therefore reports an instrumental-variable check that uses inherited proto-language gender to predict whether a respondent’s current language has sex-based grammatical gender. The relevance, or inclusion, condition is that proto-language gender strongly predicts current grammatical gender; the first-stage F-statistics in the table are large, so the instrument has substantial predictive power in these samples. The harder condition is the exclusion restriction: proto-language gender should affect current IPV, decision-making, and IPV-justification outcomes only through current gram-

⁶We merge historical institutions to DHS country-language cells using exact language-name and Glottocode matches where available, manually reviewed aliases, and transparent near matches when the D-PLACE/Ethnographic Atlas society is a defensible proxy for the DHS language label. The historical-control regressions use the same analysis-sample construction as the manuscript outcomes, but the matched sample is smaller because D-PLACE/Ethnographic Atlas coverage is strongest for African societies in our DHS country-language cells and incomplete for many non-African languages. The reported specifications include wealth quintile, education, female household head, age, age squared, country fixed effects, broad-religion fixed effects, interview-year fixed effects, and DHS survey-cluster fixed effects, with standard errors clustered by survey cluster.

matical gender. That assumption is unlikely to hold perfectly, because proto-language gender may also carry older regional, family, or gender institutions. For that reason, we present the IV estimates as a robustness check, not as a claim of clean causal identification. The new table is consistent with the baseline sign pattern for the two primary outcomes—higher reported IPV and lower household decision-making—while the IPV-justification estimate is small and not statistically precise. The exercise helps rule out simple reverse-causality stories, but it does not replace the descriptive interpretation of the main estimates. [Figure A7](#) reports the corresponding component-level IV estimates as supplemental sensitivity evidence.

Country dominance and measurement diagnostics. Another concern is that a single large DHS country or a particular measurement choice could drive the reported associations. A prior draft used an India-excluded specification to address the first concern. The refreshed analysis replaces that narrower check with a broader leave-one-country-out exercise that excludes, one at a time, the ten largest sample-size countries. [Figure A8](#) shows that the central patterns are not driven by any one of these large DHS country samples: the full-sample coefficients are +1.21 percentage points for any reported IPV, -6.25 percentage points for the household decision-making index, and +0.56 percentage points for the adjusted IPV-justification index. The remaining diagnostics examine measurement choices that could affect the language exposure. [Figure A9](#) varies the grammatical-gender assignment by formal versus semantic classification. Together with the high-confidence classification check described above, these diagnostics preserve the most consistent findings for household decision-making and reported IPV, while the smaller IPV-justification coefficients are more sensitive across components. They therefore reduce concerns about single-country leverage and exposure measurement without overstating the precision of the channel estimates.

6 Mechanisms

There are two possible channels that explain our findings. First, *cultural transmission*: gendered grammar may be part of the inherited social environment through which people learn gender roles. The second is *cognitive salience*: using a gendered language may make male and female categories more noticeable in the moment. These channels are related, and the available data cannot separate them perfectly. The goal here is to look for evidence that fits each channel and then ask which interpretation is better supported by the pattern of results.

This argument builds on two literatures. Work on linguistic relativity shows that grammatical rules can shape what speakers habitually notice, especially when a language repeatedly requires a distinction to be encoded ([Slobin, 1996](#); [Chen, 2013](#)). Experimental and psychological studies similarly find that gendered grammatical forms can affect gender categorization,

stereotypes, and social representation (Boroditsky et al., 2003; Sato et al., 2020; Stout and Dasgupta, 2011; Sczesny et al., 2016; Gygax et al., 2019). In economics, related studies connect gendered language to women’s education, labour-market outcomes, gender attitudes, and aggregate IPV patterns (Mavisakalyan, 2015; Gay et al., 2018; Jakiela and Ozier, 2020; Davis et al., 2024). Recent experimental work in economics provides a useful boundary condition for this interpretation: Balafoutas et al. (2025) randomize gender-inclusive wording in experimental instructions and find little evidence that such wording affects competitiveness, leadership, voting, or perceived inclusion in English and German laboratory samples.

6.1 Cultural Transmission: Inherited Gender Norms

The cultural-transmission channel predicts that gendered grammar should line up with gender norms learned over many years, not only with the language used in one survey conversation. In economics, social norms are often understood as shared beliefs about what behaviour others see as appropriate (Bicchieri, 2006; Krupka and Weber, 2013). We therefore first ask whether the association appears in a norm that would connect household authority to IPV. The DHS asks whether a husband is justified in hitting or beating his wife in specific situations: if she goes out without telling him, neglects the children, argues, refuses sex, does not cook properly, is suspected of infidelity, or is disrespectful. These questions are useful because they are not about violence in the abstract. They are about whether a husband may discipline his wife when she violates expected rules of movement, sexuality, obedience, or domestic work. Prior work uses wife-beating justification as a measure of IPV-related norms and links those norms to IPV risk (Sardinha and Nájera Catalán, 2018; Cools and Kotsadam, 2017; Ghoshal et al., 2022; Gunarathne et al., 2023).

Table 3 shows that speaking a sex-based gendered language is associated with a 0.6 percentage point increase in the adjusted IPV-related belief index. The compact adjusted indicators move in the same direction: the Any adjusted measure rises by 0.6 percentage points, while the None adjusted measure falls by 0.6 percentage points. The component columns are interpreted according to their displayed coding. Gendered language is associated with greater acceptance of wife beating when a wife refuses sex (+0.8 percentage points) or does not cook properly (+0.7 percentage points); the arguing column is coded as rejection of justification and is therefore read separately.

Figure 4 then asks whether this is really a gender-specific pattern. If gendered grammar were simply standing in for general conservatism or general comfort with violence, we would expect similar associations for less gender-specific attitudes. Instead, the pattern is stronger for measures that directly concern gender hierarchy inside intimate relationships. This shows that the main outcomes line up with a specific norm about marital authority, not just with a broad

tendency to accept violence.

The native-language comparison gives a second clue. DHS waves that record both native language and interview language allow us to ask whether long-run language exposure is more informative than the language used in the survey. [Figure 5](#) shows that native-language grammatical gender is more strongly associated with the primary outcomes. For household decision-making, both native-language and interview-language grammatical gender are negatively associated with participation in healthcare, major purchases, and family visits, but the native-language coefficients are larger. Reported IPV also points more strongly to native language: native-language grammatical gender is positively associated with physical IPV and any IPV, while interview-language coefficients are smaller and more mixed. This pattern fits the idea that household bargaining and IPV are shaped by family, community, and marital norms learned over many years.

6.2 Cognitive Salience: Current Language Context

The cognitive-salience channel makes a different prediction. If the immediate language context is central, then the gender structure of the survey interview language should be strongly related to the outcomes. This is plausible because experiments show that language can change which social categories feel salient in a given moment, including among bilingual respondents ([Pérez and Tavits, 2019](#)). It is also the channel most closely related to short-run language interventions. For example, [Balafoutas et al. \(2025\)](#) randomize gender-inclusive wording in experimental instructions and find little evidence that the immediate wording changes competitiveness, willingness to lead, voting for leaders, or feelings of inclusion. That result is relevant for policy interpretation because it suggests that inclusive wording may affect some representational outcomes without necessarily changing costly economic behavior in a single exposure ([Sczesny et al., 2016](#); [Tavits and Pérez, 2019](#); [Balafoutas et al., 2025](#)).

The evidence for this channel is weaker. In [Figure 5](#), interview-language grammatical gender is associated with some household decision-making outcomes, but the coefficients are generally smaller than the native-language coefficients and are more mixed for reported IPV. This makes current language context a less convincing explanation for the primary outcomes. The result is consistent with the experimental evidence above: a language cue encountered in the moment may be too weak to move high-stakes economic behaviour by itself, even if language matters through slower socialization. It may still matter for stated attitudes: IPV-justification results are more mixed in the restricted native-language samples, and survey language could influence how respondents answer questions about gender roles in the moment.

7 Conclusion

This paper examines whether sex-based grammatical gender is associated with two central development outcomes for married women: safety from intimate partner violence and agency in household decisions. Across 126 DHS rounds in 53 countries, women speaking languages with sex-based grammatical gender systems report more IPV experience and less participation in decisions about healthcare, purchases, and family visits. The IPV estimates are modest in percentage-point terms, but they concern a serious harm that is difficult to measure and often underreported. The decision-making estimates point to meaningful differences in everyday household agency.

The mechanism evidence points more strongly to cultural transmission than to short-run cognitive salience. Women speaking gendered languages are more likely to accept wife beating in scenarios tied to domestic compliance, sexual availability, and disagreement with a husband. The specificity checks and native-versus-interview-language comparison suggest that the association is strongest for inherited exposure and gender-hierarchical content, not for general violence attitudes or arbitrary linguistic structure.

The policy implication is not that grammar of a language is a simple lever for change. Rather, language is part of the environment in which gender-equality policies operate. Gender-inclusive language guidance in public institutions is built on the premise that official communication should avoid masculine defaults and stereotype-reinforcing formulations, while research on gender-fair language shows that linguistic form can affect representation, salience, and stereotypes (United Nations, 2026; Bigler and Leaper, 2015; Sczesny et al., 2016). Our results support treating such language-aware practices as complements to deeper development policies, not as substitutes for IPV prevention, schooling, income support, legal protection, or improvements in women’s bargaining power. This caution is consistent with recent experimental evidence showing limited short-run effects of one-time inclusive wording on economic behavior (Balafoutas et al., 2025). The attenuation by schooling and the urban-rural patterns suggest that formal education, exposure to alternative norms, and stronger public institutions can weaken the relationship between linguistic structure and adverse gender outcomes. Language-of-instruction choices, bilingual education, public communication, and curricula that help students reflect on gendered linguistic forms may therefore complement legal and economic reforms (Sczesny et al., 2016; Bigler and Leaper, 2015).

Several limitations remain. The estimates are descriptive associations, not strict causal effects. Fixed effects, placebo tests, ethnicity controls, leave-one-country-out diagnostics, historical-institution controls, and the proto-language IV check reduce some concerns, but unobserved linguistic-community characteristics may still matter. We also observe only one or two language

measures per respondent, while multilingual women may use different languages across settings. Future data with fuller language histories, partner beliefs, community norms, and household dynamics would allow sharper tests of the channels suggested here.

References

- Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. *Quarterly Journal of Economics*, 115(3):715–753.
- Alesina, A., Brioschi, B., and La Ferrara, E. (2021). Violence against women: A cross-cultural analysis for africa. *Economica*, 88(349):70–104.
- Alesina, A., Giuliano, P., and Nunn, N. (2013). On the origins of gender roles: Women and the plough. *Quarterly Journal of Economics*, 128(2):469–530.
- Annan, J., Donald, A., Goldstein, M., Martinez, P., and Koolwal, G. (2021). Taking power: Women’s empowerment and household well-being in sub-Saharan Africa. *World Development*, 140:105292.
- Balafoutas, L., Fornwagner, H., Hauser, E., and Hauser, O. P. (2025). Gender-inclusive language and economic decision-making. *Management Science*. Published online December 19, 2025.
- Bengesai, A. V. and Khan, H. T. A. (2023). Exploring the association between attitudes towards wife beating and intimate partner violence using a dyadic approach in three sub-saharan african countries. *BMJ Open*, 13:e062977.
- Benjamin, D. J., Choi, J. J., and Strickland, A. J. (2010). Social identity and preferences. *American Economic Review*, 100(4):1913–1928.
- Bicchieri, C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press, New York.
- Bigler, R. S. and Leaper, C. (2015). Gendered language: Psychological principles, evolving practices, and inclusive policies. *Policy Insights from the Behavioral and Brain Sciences*, 2(1):187–194.
- Bisin, A. and Verdier, T. (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory*, 97(2):298–319.
- Boroditsky, L., Schmidt, L. A., and Phillips, W. (2003). Sex, syntax, and semantics. In Gentner, D. and Goldin-Meadow, S., editors, *Language in Mind: Advances in the Study of Language and Cognition*, pages 61–80. MIT Press, Cambridge, MA.
- Chen, M. K. (2013). The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. *American Economic Review*, 103(2):690–731.
- Cohen, A., Karelitz, T., Kricheli-Katz, T., Pumpian, S., and Regev, T. (2023). Gender-neutral language and gender disparities. Working Paper 31400, National Bureau of Economic Research.

- Cohn, A., Fehr, E., and Marechal, M. A. (2014). Business culture and dishonesty in the banking industry. *Nature*, 516:86–89.
- Cools, S. and Kotsadam, A. (2017). Resources and intimate partner violence in sub-saharan africa. *World Development*, 95:211–230.
- Corbett, G. (2013a). Number of genders. In: Dryer, matthew s. & haspelmath, martin (eds.) WALS online (v2020.4) [30a].
- Corbett, G. (2013b). Sex-based and non-sex-based gender systems. In: Dryer, matthew s. & haspelmath, martin (eds.) WALS online (v2020.4) [31a].
- Corbett, G. (2013c). Systems of gender assignment. In: Dryer, matthew s. & haspelmath, martin (eds.) WALS online (v2020.4) [32a].
- Cubelli, R., Paolieri, D., Lotto, L., and Job, R. (2011). The effect of grammatical gender on object categorization. *Journal of Experimental Psychology*, 37(2):449–460.
- D-PLACE (2026). D-PLACE aggregated CLDF dataset, version 3.2.0. Accessed July 23, 2026. Canonical DOI: 10.5281/zenodo.3935419.
- Davis, L., Mavisakalyan, A., and Weber, C. (2024). Gendered language and gendered violence. *Journal of Comparative Economics*.
- Davis, L. and Reynolds, M. (2018). Gendered language and the educational gender gap. *Economics Letters*, 168:46–48.
- Dhar, D., Jain, T., and Jayachandran, S. (2022). Reshaping adolescents’ gender attitudes: Evidence from a school-based experiment in india. *American Economic Review*, 112(3):899–927.
- Duflo, E. (2012). Women empowerment and economic development. *Journal of Economic Literature*, 50(4):1051–1079.
- Gabriel, U., Gyax, P. M., and Kuhn, E. A. (2018). Neutralising linguistic sexism: Promising but cumbersome? *Group Processes & Intergroup Relations*, 21(5):844–858.
- Gaucher, D., Friesen, J., and Kay, A. C. (2011). Evidence that gendered wording in job advertisements exists and sustains gender inequality. *Journal of Personality and Social Psychology*, 101(1):109–128.
- Gay, V., Hicks, D. L., Santacreu-Vasut, E., and Shoham, A. (2018). Decomposing culture: An analysis of gender, language, and labor supply in the household. *Review of Economics of the Household*, 16(4):879–909.

- Ghoshal, R., Douard, A.-C., Sikder, S., Roy, N., and Saulnier, D. D. (2022). Risk and protective factors for IPV in low- and middle-income countries: A systematic review. *Journal of Aggression, Maltreatment & Trauma*, 32(4):505–522.
- Giuliano, P. (2020). Gender and culture. Working Paper 27725, National Bureau of Economic Research.
- Guiso, L., Sapienza, P., and Zingales, L. (2006). Does culture affect economic outcomes? *Journal of Economic Perspectives*, 20(2):23–48.
- Gunarathne, L., Bhowmik, J., Apputhurai, P., and Nedeljkovic, M. (2023). Factors and consequences associated with intimate partner violence against women in low- and middle-income countries: A systematic review. *PLOS ONE*, 18(11):e0293295.
- Gygax, P., Elmiger, D., Zufferey, S., Garnham, A., Sczesny, S., von Stockhausen, L., Braun, F., and Oakhill, J. (2019). A language index of grammatical gender dimensions to study the impact of grammatical gender on the way we perceive women and men. *Frontiers in Psychology*, 10:1604.
- Haynie, H. J., Blasi, D. E., Skirg ard, H., Greenhill, S. J., Atkinson, Q. D., and Gray, R. D. (2023). Grambank’s typological advances support computational research on diverse languages. In *Proceedings of the 5th Workshop on Research in Computational Linguistic Typology and Multilingual NLP*, pages 147–149.
- Heath, R. (2014). Women’s access to labor market opportunities, control of household resources, and domestic violence: Evidence from Bangladesh. *World Development*, 57:32–46.
- Hicks, D. L., Santacreu-Vasut, E., and Shoham, A. (2015). Does mother tongue make for women’s work? linguistics, household labor, and gender identity. *Journal of Economic Behavior & Organization*, 110:19–44.
- Hidrobo, M. and Fernald, L. (2013). Cash transfers and domestic violence. *Journal of Health Economics*, 32(1):304–319.
- Hoff, K. and Pandey, P. (2006). Discrimination, social identity, and durable inequalities. *American Economic Review*, 96(2):206–211.
- Jakiela, P. and Ozier, O. (2020). Gendered language. Discussion Paper 13126, IZA Discussion Papers.
- Jayachandran, S. (2015). The roots of gender inequality in developing countries. *Annual Review of Economics*, 7:63–88.

- Jayachandran, S. (2021). Social norms as a barrier to women’s employment in developing countries. *IMF Economic Review*, 69(3):576–595.
- Jayachandran, S. and Voena, A. (2026). Women’s power in the household. *Journal of Economic Literature*, 64(2):447–497.
- Jensen, R. and Oster, E. (2009). The power of TV: Cable television and women’s status in india. *Quarterly Journal of Economics*, 124(3):1057–1094.
- Kabeer, N. (1999). Resources, agency, achievements: Reflections on the measurement of women’s empowerment. *Development and Change*, 30(3):435–464.
- Kadengye, D. T., Izudi, J., Kemigisha, E., and Kiwuwa-Muyingo, S. (2023). Effect of justification of wife-beating on experiences of intimate partner violence among men and women in uganda: A propensity-score matched analysis of the 2016 demographic health survey data. *PLOS ONE*, 18:e0276025.
- Keller, T., Kiss, H. J., and Szakál, P. (2024). Endogenous language use and patience. *Journal of Economic Behavior & Organization*, 220:792–812.
- Kirby, K. R., Gray, R. D., Greenhill, S. J., Jordan, F. M., Gomes-Ng, S., Bibiko, H.-J., Blasi, D. E., Botero, C. A., Bown, C., Ember, C. R., Leehr, D., Low, B. S., McCarter, J., Divale, W., and Gavin, M. C. (2016). D-PLACE: A global database of cultural, linguistic and environmental diversity. *PLOS ONE*, 11(7):e0158391.
- Kishor, S. and Subaiya, L. (2008). Understanding women’s empowerment: A comparative analysis of Demographic and Health Surveys (DHS) data. *DHS Comparative Reports*, 20.
- Kricheli-Katz, T. and Regev, T. (2021). The effect of language on performance: Do gendered languages fail women in maths? *npj Science of Learning*, 6:9.
- Krupka, E. L. and Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3):495–524.
- Lange, E. and von Stockhausen, L. (2025). What affects people’s use of and attitudes toward gender-inclusive language? a literature review. *Frontiers in Psychology*, 16:1657753.
- Lian, Z., Lien, D., and Sun, J. (2026). The linguistic driver of divorce. *Journal of Economic Behavior & Organization*, 241:107350.
- Lindqvist, A., Renström, E. A., and Gustafsson Sendén, M. (2019). Reducing a male bias in language? establishing the efficiency of three different gender-fair language strategies. *Sex Roles*, 81(1–2):109–117.

- Lundberg, S. and Pollak, R. A. (1993). Separate spheres bargaining and the marriage market. *Journal of Political Economy*, 101(6):988–1010.
- Mavisakalyan, A. (2015). Gender in language and gender in employment. *Oxford Development Studies*, 43(4):403–424.
- Michaelis, S. M., Maurer, P., Haspelmath, M., and Huber, M., editors (2013). *Atlas of Pidgin and Creole Language Structures Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig.
- Murdock, G. P. (1967). Ethnographic atlas: A summary. *Ethnology*, 6(2):109–236.
- Nabaggala, M. S., Reddy, T., and Manda, S. (2021). Effects of rural–urban residence and education on intimate partner violence among women in sub-saharan africa: A meta-analysis of health survey data. *BMC Women’s Health*, 21(1):149.
- Pérez, E. O. and Tavits, M. (2019). Language influences public attitudes toward gender equality. *Journal of Politics*, 81(1):81–93.
- Samuel, S., Cole, G., and Eacott, M. J. (2019). Grammatical gender and linguistic relativity: A systematic review. *Psychonomic Bulletin & Review*, 26:1767–1786.
- Sardinha, L. and Nájera Catalán, H. E. (2018). Attitudes towards domestic violence in 49 low- and middle-income countries: A gendered analysis of prevalence and country-level correlates. *PLOS ONE*, 13(10):e0206101.
- Sato, S., Casaponsa, A., and Athanasopoulos, P. (2020). Flexing gender perception: Brain potentials reveal the cognitive permeability of grammatical information. *Cognitive Science*, 44(9):e12884.
- Sczesny, S., Formanowicz, M., and Moser, F. (2016). Can gender-fair language reduce gender stereotyping and discrimination? *Frontiers in Psychology*, 7:25.
- Shah, M. and Barski, L. (2026). Intimate partner violence in low- and middle-income countries: Insights from economic research. *Journal of Economic Literature*, 64(2).
- Shayo, M. (2020). Social identity and economic policy. *Annual Review of Economics*, 12:355–389.
- Skirgård, H., Haynie, H. J., Blasi, D. E., Hammarström, H., Collins, J., Latache, J. J., Lesage, J., Weber, T., Witzlack-Makarevich, A., Passmore, S., et al. (2023). Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss. *Science Advances*, 9(16):eadg6175.

- Slobin, D. (1996). From ‘thought and language’ to ‘thinking for speaking’. In Gumperz, J. J. and Levinson, S. C., editors, *Rethinking Linguistic Relativity*, pages 70–96. Cambridge University Press, Cambridge.
- Stout, J. G. and Dasgupta, N. (2011). When he doesn’t mean you: Gender-exclusive language as ostracism. *Personality and Social Psychology Bulletin*, 37(6):757–769.
- Tavits, M. and Pérez, E. O. (2019). Language influences mass opinion toward gender and LGBT equality. *Proceedings of the National Academy of Sciences*, 116(34):16781–16786.
- The DHS Program (2026a). Demographic and health survey (DHS). Accessed July 21, 2026. <https://www.dhsprogram.com/What-We-Do/Survey-Types/DHS.cfm>.
- The DHS Program (2026b). The DHS program application programming interface. Accessed July 21, 2026. <https://api.dhsprogram.com/>.
- The DHS Program (2026c). Guide to DHS statistics: Domestic violence. Accessed July 21, 2026. https://www.dhsprogram.com/Data/Guide-to-DHS-Statistics/17_Domestic_Violence.htm.
- Tran, T. D., Nguyen, H., and Fisher, J. (2016). Attitudes towards intimate partner violence against women among women and men in 39 low- and middle-income countries. *PLOS ONE*, 11(11):e0167438.
- Tur-Prats, A. (2019). Family types and intimate partner violence: A historical perspective. *Review of Economics and Statistics*, 101(5):878–891.
- United Nations (2026). Gender-inclusive language. Accessed July 24, 2026. <https://www.un.org/en/gender-inclusive-language/>.
- Van der Velde, L., Tyrowicz, J., and Siwinska, J. (2015). Language and (the estimates of) the gender wage gap. *Economics Letters*, 136:165–170.
- Vervecken, D., Hannover, B., and Wolter, I. (2013). Changing (s)expectations: How gender fair job descriptions impact children’s perceptions and interest regarding traditionally male occupations. *Journal of Vocational Behavior*, 82(3):208–220.
- Whorf, B. L. (1956). *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*. MIT Press, Cambridge, MA.
- World Bank (2026a). World bank country and lending groups. FY27 income classifications, effective July 1, 2026. Accessed July 21, 2026. <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>.

World Bank (2026b). World development indicators: Population, total (SP.POP.TOTL) and country API. API data accessed July 21, 2026. <https://api.worldbank.org/v2/country/all/indicator/SP.POP.TOTL>.

Table 1: Descriptive Statistics by Sex-Based Grammatical Gender

Variable	Obs.	Mean	S.D.	Mean, gendered	Mean, non-gendered
Panel A: Language exposure and sample structure					
District/region has more than 2 languages	1,531,579	0.742	0.438	0.794	0.698
Sex-based grammatical gender	1,531,579	0.459	0.498	1.000	0.000
Panel B: Reported IPV experience					
Physical IPV	579,611	0.052	0.223	0.049	0.056
Sexual IPV	574,916	0.021	0.144	0.015	0.027
Emotional IPV	572,813	0.052	0.221	0.045	0.058
Any reported IPV	580,200	0.084	0.278	0.074	0.094
Panel C: Household decision-making					
Healthcare decision	1,061,923	0.657	0.475	0.649	0.664
Daily purchases	312,385	0.615	0.487	0.627	0.606
Major purchases	1,053,602	0.612	0.487	0.619	0.607
Visit family	1,053,235	0.692	0.462	0.680	0.703
Panel D: IPV-justification measures					
IPV justified: goes out	1,512,924	0.255	0.436	0.232	0.274
IPV justified: neglects child	1,518,817	0.294	0.456	0.272	0.313
IPV justified: argues with husband	1,508,853	0.245	0.430	0.234	0.255
IPV justified: refuses sex	1,468,575	0.186	0.389	0.166	0.202
IPV justified: does not cook	1,513,383	0.143	0.350	0.149	0.137
Panel E: Baseline controls					
Wealth quintile	1,531,579	3.130	1.422	3.158	3.106
Highest education level	1,531,579	1.352	0.965	1.340	1.362
Female household head	1,531,579	0.231	0.422	0.200	0.258
Age	1,531,579	29.449	9.686	29.480	29.423

Notes: The table reports descriptive statistics for the analysis samples used in Tables 2 and 3. Reported-IPV rows use the corresponding Table 2 Panel A regression sample; household-decision rows use the corresponding Panel B regression sample; IPV-justification rows, language exposure, sample structure, and controls use the Table 3 baseline sample. Obs. is the row-specific sample size. Group means split the same row-specific sample by whether the respondent’s assigned primary language has sex-based grammatical gender. Household-decision variables equal one when the respondent participates alone or jointly; IPV-justification variables equal one when wife beating is justified in the listed scenario.

[\[↑ back to text, Data and measurement\]](#)

Table 2: Sex-Based Grammatical Gender and Primary Outcomes

Panel A: Reported IPV Experience					
	Any IPV	Physical	Sexual	Emotional	
Gendered language	0.012*** (0.002)	0.012*** (0.002)	0.003** (0.001)	0.004** (0.002)	
Observations	580,200	579,611	574,916	572,813	
Mean dependent variable	0.084	0.052	0.021	0.052	
Panel B: Household Decision-Making					
	Index	Healthcare	Daily purchases	Major purchases	Visit family
Gendered language	-0.062*** (0.003)	-0.066*** (0.003)	-0.046*** (0.005)	-0.053*** (0.003)	-0.065*** (0.003)
Observations	1,062,684	1,061,923	312,385	1,053,602	1,053,235
Mean dependent variable	0.657	0.657	0.615	0.612	0.692

Notes: The table reports coefficients on sex-based grammatical gender from the baseline linear probability specification. Panel A uses the DHS domestic-violence-module sample and measures reported IPV experience. Panel B uses the broader available household-decision sample and measures whether the respondent participates alone or jointly in each decision. All specifications include wealth quintile, highest educational level, female household head, respondent age, and respondent age squared, with country, broad religion, interview-year, and survey-cluster fixed effects. Standard errors are clustered by survey cluster and appear in parentheses. The stay-family decision item is omitted because it has no available observations in the current analysis file.

[\[↑ back to text, Results\]](#)

Table 3: IPV-Related Beliefs and Sex-Based Grammatical Gender

	Adjusted index	Any adjusted	None adjusted	Rejects: argues	Justified: refuses sex	Justified: does not cook
Gendered language	0.006*** (0.001)	0.006*** (0.001)	-0.006*** (0.001)	0.005** (0.003)	0.008*** (0.002)	0.007*** (0.002)
Observations	1,531,579	1,531,579	1,531,579	1,508,853	1,468,575	1,513,383
Mean dependent variable	0.496	0.968	0.032	0.755	0.186	0.143

Notes: The table reports coefficients on sex-based grammatical gender for compact IPV-related belief measures used as channel evidence. The adjusted index is the respondent-level mean of available wife-beating items after item-specific coding adjustments; Any adjusted and None adjusted indicate whether any or no adjusted item equals one. Because the adjusted component coding is item-specific, component columns are labeled by their actual coding: Rejects: argues equals one if the respondent says wife beating is not justified when a wife argues with her husband, while Justified: refuses sex and Justified: does not cook equal one if the respondent accepts wife beating in those scenarios. These estimates do not test mediation from women’s own beliefs to their victimization; they proxy the gender-hierarchical normative environment in which IPV may be more socially tolerated. Specifications include the baseline controls and country, broad religion, interview-year, and survey-cluster fixed effects, with standard errors clustered by survey cluster.

[\[↑ back to text, Mechanisms\]](#)

Table 4: Historical-Institution Controls

Specification	Any reported IPV	Decision index	IPV-justification index
Baseline matched sample	0.009** (0.004)	-0.230*** (0.007)	0.007*** (0.002)
Add plough	0.009** (0.004)	-0.235*** (0.007)	0.006** (0.002)
Add descent	0.005 (0.004)	-0.217*** (0.007)	0.001 (0.002)
Add residence	0.010*** (0.004)	-0.233*** (0.007)	0.008*** (0.002)
Bride/dowry controls	0.009*** (0.004)	-0.231*** (0.007)	0.006*** (0.002)
Joint extended controls	0.002 (0.004)	-0.234*** (0.007)	0.000 (0.002)

Notes: Each entry is the coefficient on sex-based grammatical gender from a separate regression that adds the historical-institution controls listed in the row. Standard errors clustered by DHS survey cluster appear in parentheses. The baseline matched-sample row uses the manuscript controls and fixed effects but restricts the sample to observations with nonmissing historical-institution measures. The remaining rows add ancestral plough use; patrilineal and matrilineal descent; patrilocal and matrilineal residence; bride price or bride wealth and dowry; or all historical controls jointly. All specifications include wealth quintile, education, female household head, age, age squared, country fixed effects, broad-religion fixed effects, interview-year fixed effects, and DHS survey-cluster fixed effects. Outcome-specific samples are 246,729 observations for any reported IPV, 446,880 for the decision index, and 629,826 for the IPV-justification index. The historical variables come from D-PLACE/Ethnographic Atlas country-language matches. Stability of the coefficient after adding historical controls indicates that the language association is not fully absorbed by inherited institutions; attenuation indicates that grammatical gender is partly bundled with broader historical gender institutions.

[\[↑ back to text, Robustness checks\]](#)

Table 5: Instrumental-Variable Estimates Using Proto-Language Gender

	Any IPV	Decision index	IPV-justification index
IV gendered language	0.093*** (0.012)	-0.194*** (0.015)	0.010 (0.006)
Observations	403,868	727,352	1,029,697
First-stage F	1079.37	2389.36	2303.67

Notes: The table reports instrumental-variable estimates in which inherited proto-language gender structure instruments for current sex-based grammatical gender. Columns follow the manuscript outcome hierarchy: reported IPV experience, household decision-making, and IPV-justification channel evidence. The first-stage F-statistic is reported for each column. These estimates are interpreted as sensitivity evidence against simple reverse causality, not as central causal magnitudes, because proto-language gender may also proxy for other persistent cultural or regional characteristics.

[\[↑ back to text, Robustness checks\]](#)

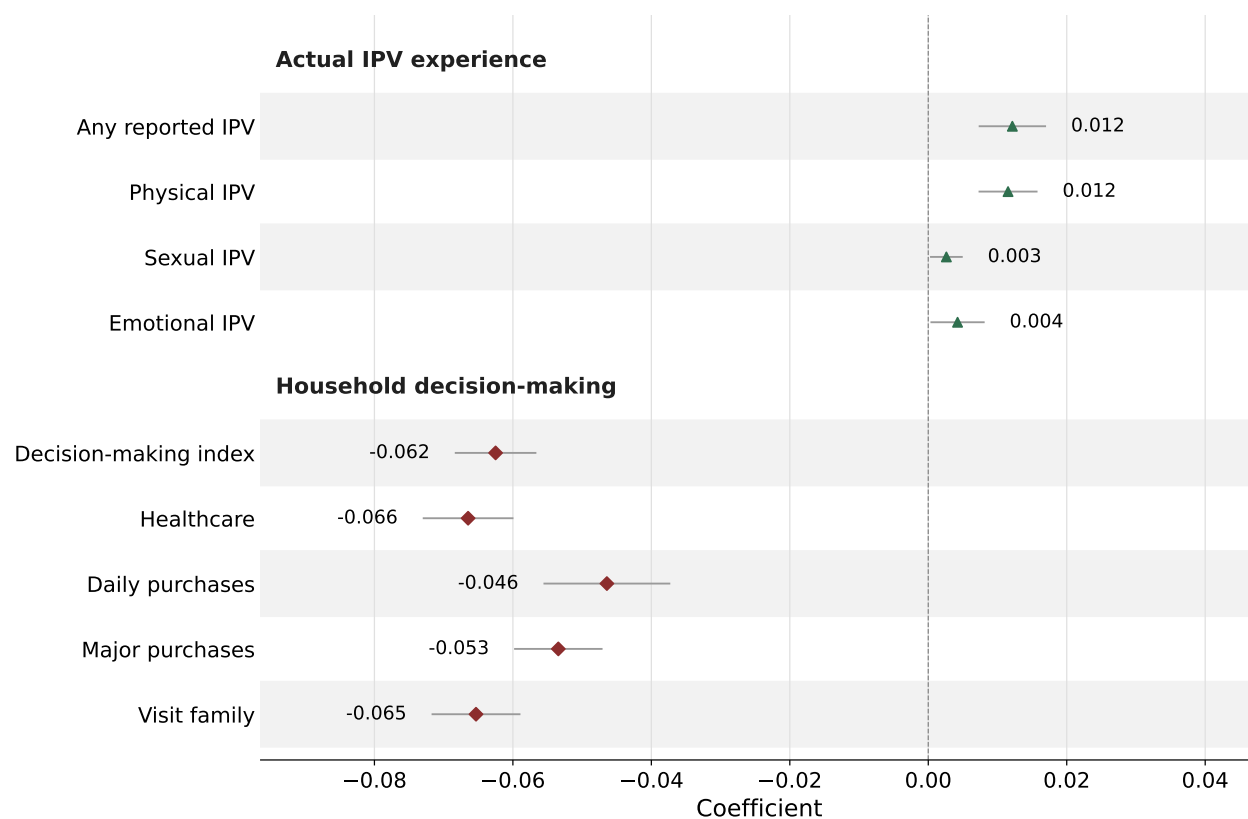


Figure 1: Primary Outcome Estimates

Notes: The figure plots coefficients on whether the respondent's assigned primary language has sex-based grammatical gender for the two primary outcome domains: reported IPV experience and household decision-making. Reported-IPV outcomes use the DHS domestic-violence-module sample; household-decision outcomes use the broader available decision-making sample. Points show coefficients and horizontal bars show 95 percent confidence intervals from the baseline specification. Specifications include wealth quintile, highest educational level, female household head, respondent age, and respondent age squared, with country, broad religion, interview-year, and survey-cluster fixed effects. Standard errors are clustered by survey cluster. The vertical reference line marks zero.

[\[↑ back to text, Results\]](#)

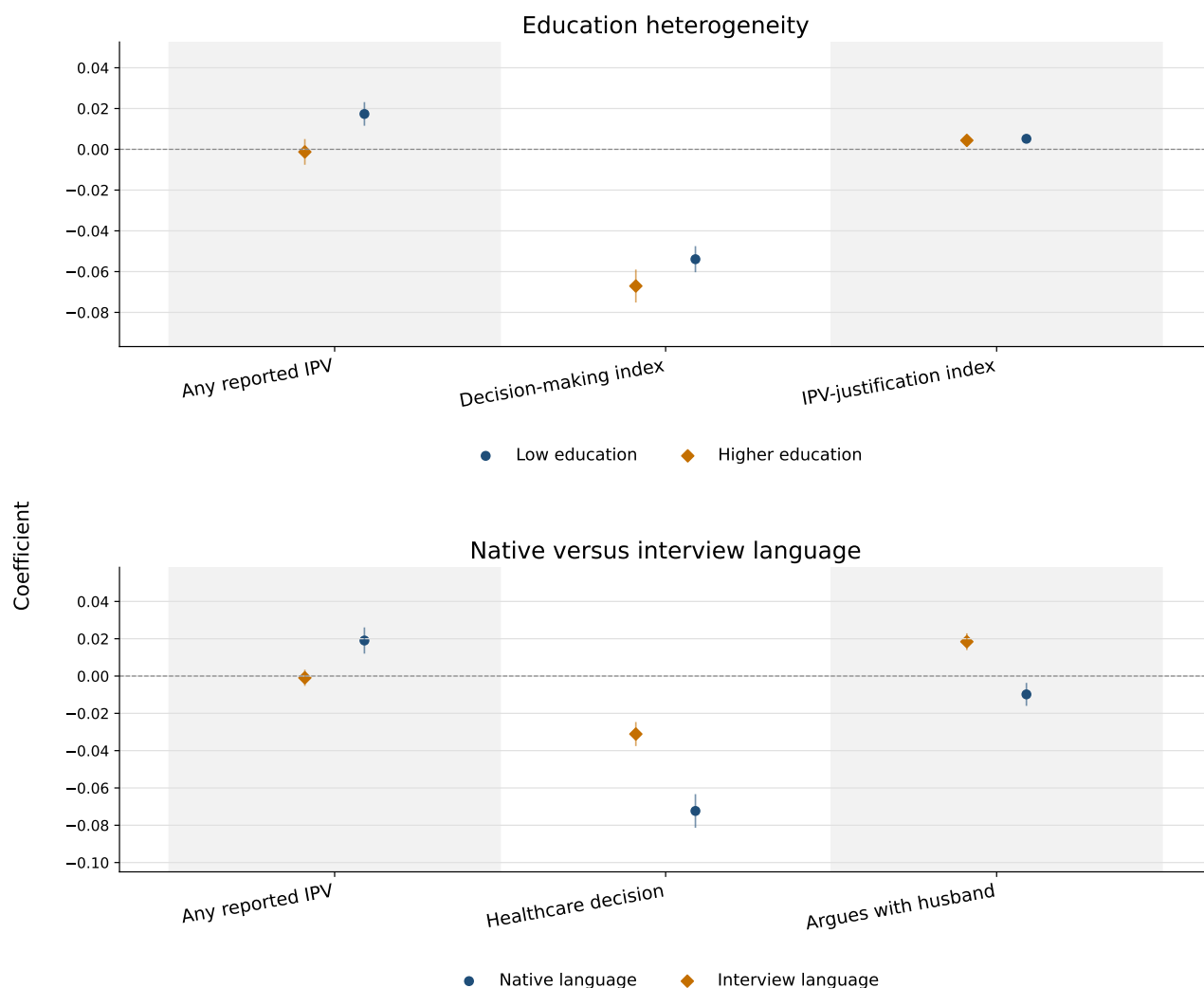


Figure 2: Education Heterogeneity and Language Context

Notes: Panel A reports education heterogeneity for summary outcomes. Panel B compares native-language and interview-language grammatical-gender measures for representative outcomes. Points show coefficients and horizontal bars show 95 percent confidence intervals. Specifications retain the baseline controls and fixed-effect structure where supported by the relevant sample. The figure is interpreted as descriptive heterogeneity and suggestive evidence on channel timing, not as a clean decomposition of cognitive priming and cultural transmission.

[\[↑ back to text, Heterogeneity\]](#)

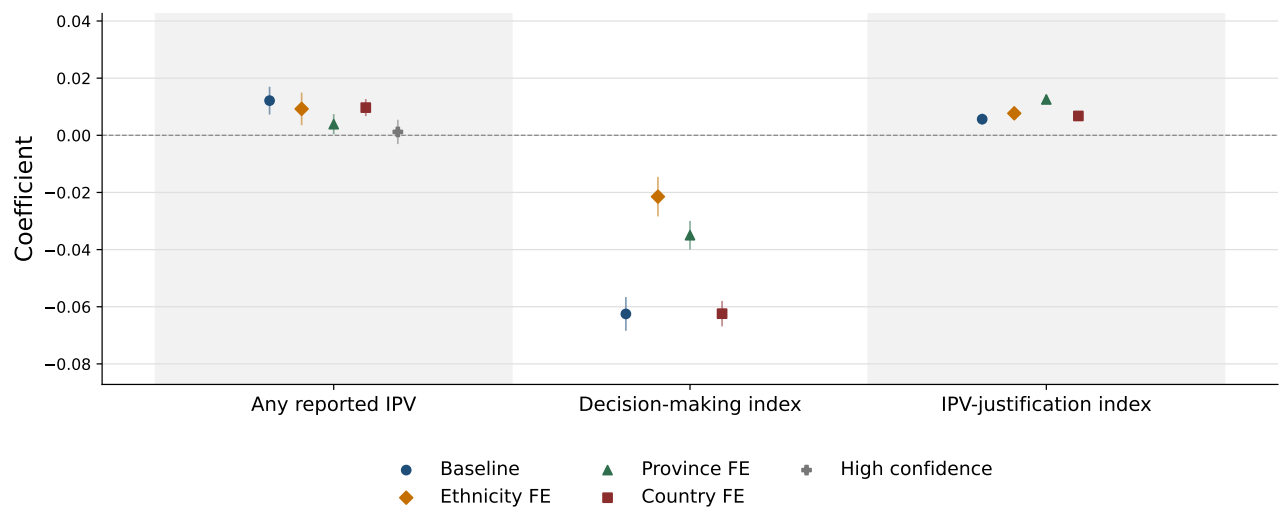


Figure 3: Fixed-Effect and Source-Quality Robustness

Notes: The figure summarizes baseline, ethnicity fixed-effect, province fixed-effect, country fixed-effect, and high-confidence language-classification specifications for the main summary outcomes and the IPV-justification channel index. Points show coefficients on sex-based grammatical gender and horizontal bars show 95 percent confidence intervals. Specifications are organized around reported IPV experience, household decision-making, and IPV-justification channel evidence. The figure assesses whether the main hierarchy is sensitive to ethnicity controls, geographic fixed-effect structure, or source-quality restrictions.

[\[↑ back to text, Robustness checks\]](#)

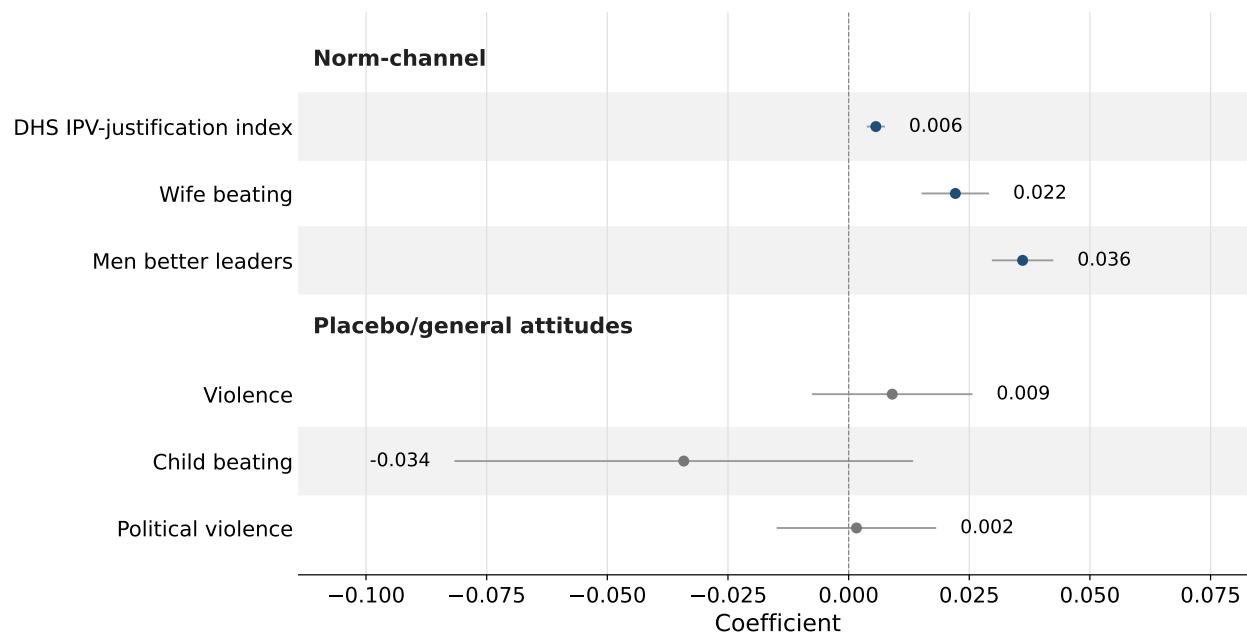


Figure 4: Norm-Specificity Evidence

Notes: The figure contrasts gender-hierarchical norm measures with placebo or less gender-specific outcomes. Points show coefficients on sex-based grammatical gender and horizontal bars show 95 percent confidence intervals. The outcome order places reported IPV experience and household decision-making before IPV-justification measures where those outcomes appear. The figure shows whether the association is strongest for gender-hierarchical content rather than for general violence attitudes or broad conservatism.

[\[↑ back to text, Mechanisms\]](#)

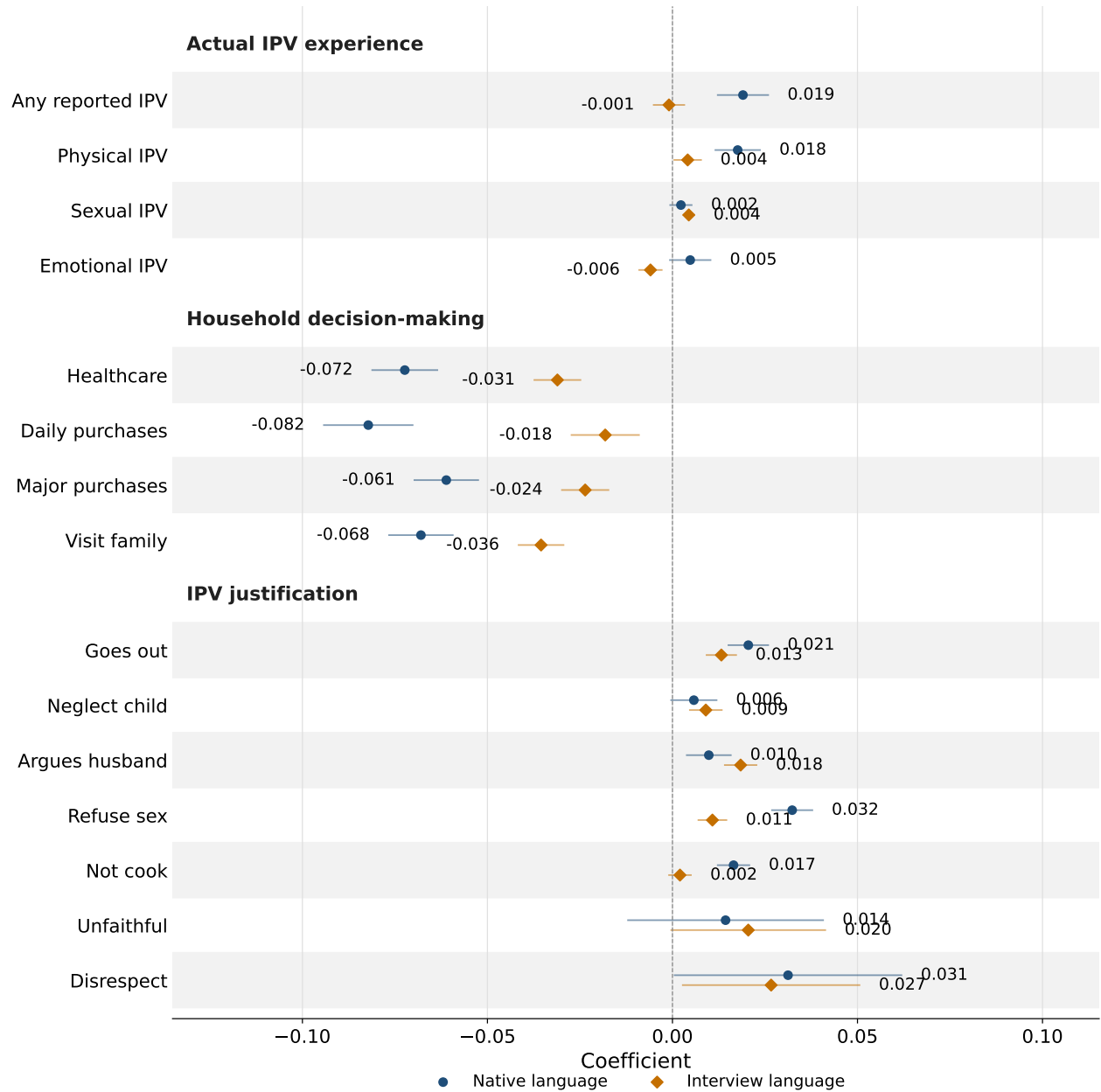


Figure 5: Native Versus Interview Language

Notes: The figure compares coefficients based on sex-based grammatical gender in the respondent's native language with coefficients based on sex-based grammatical gender in the interview language. Points show coefficients and horizontal bars show 95 percent confidence intervals. Specifications include the baseline controls and fixed effects where the native-language and interview-language measures are observed. The comparison is suggestive channel evidence: native-language coefficients speak more directly to inherited cultural transmission, while interview-language coefficients speak more directly to contemporaneous linguistic context.

[\[↑ back to text, Mechanisms\]](#)

Appendix A. Supplemental Tables and Figures

Table A1: Same-Sample Estimates in the DHS Domestic-Violence Module

Panel A: Reported IPV Experience					
	Any IPV	Physical	Sexual	Emotional	
Gendered language	0.012*** (0.002)	0.012*** (0.002)	0.003** (0.001)	0.004** (0.002)	
Observations	580,200	579,611	574,916	572,813	
Panel B: Household Decision-Making					
	Index	Healthcare	Daily purchases	Major purchases	Visit family
Gendered language	-0.078*** (0.004)	-0.087*** (0.004)	-0.043*** (0.007)	-0.065*** (0.004)	-0.083*** (0.004)
Observations	522,075	521,904	147,672	521,768	521,730
Panel C: IPV-Related Beliefs					
	Index	Argues	Refuses sex	Does not cook	
Gendered language	0.000 (0.001)	0.009*** (0.003)	0.024*** (0.003)	0.012*** (0.003)	
Observations	578,010	573,426	557,882	574,205	

Notes: The table repeats the primary outcome hierarchy after restricting each outcome to respondents in the DHS domestic-violence-module pool where the corresponding outcome is available. Panel A reports IPV experience; Panel B reports household decision-making participation; Panel C reports compact IPV-related belief measures. The exercise addresses whether differences between the domestic-violence-module sample and the broader decision-making or belief samples drive the main pattern. It is not a causal mediation estimate. Specifications use the baseline controls and fixed effects, and standard errors are clustered by survey cluster.

[\[↑ back to text, Robustness checks\]](#)

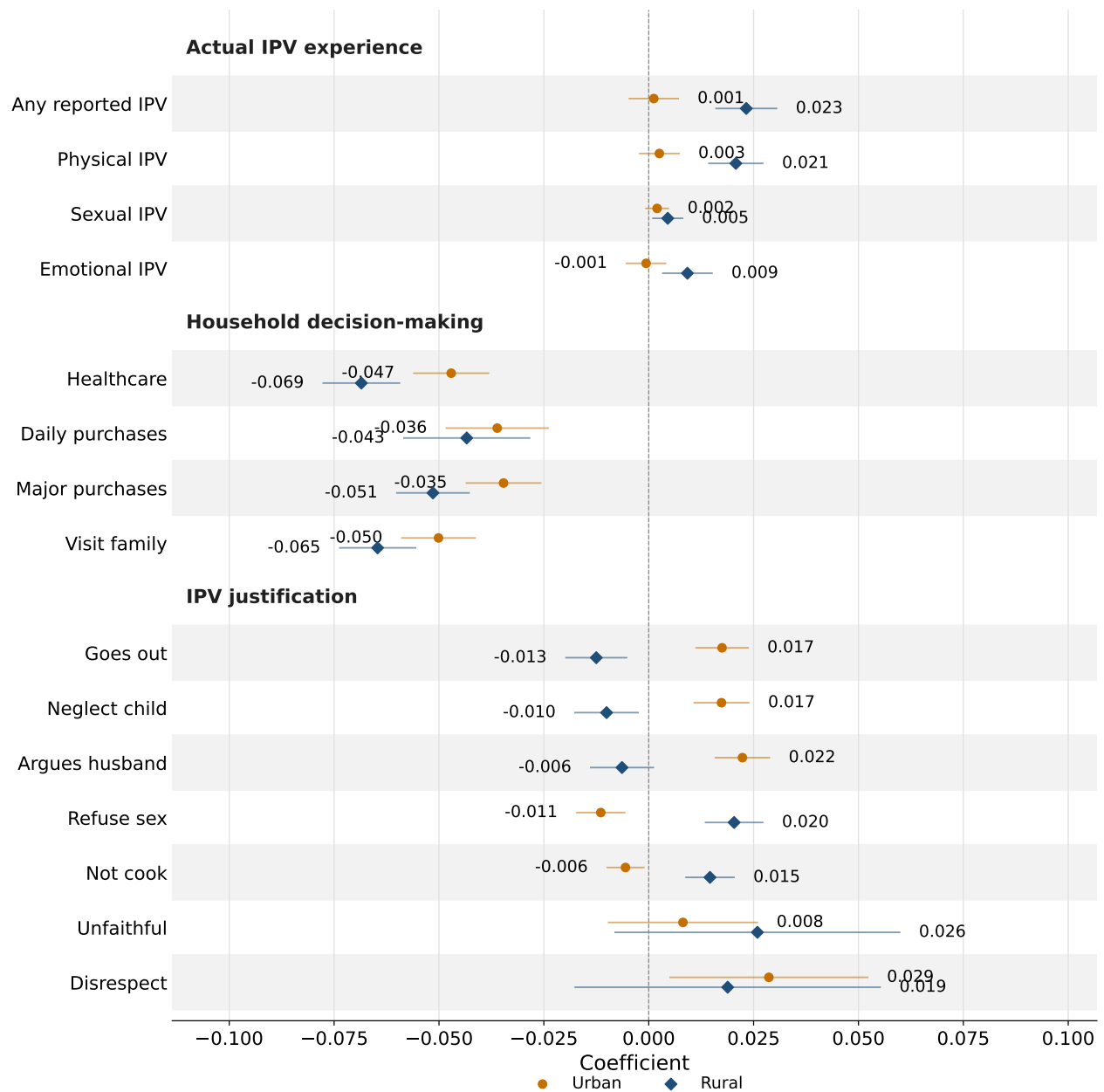


Figure A1: Urban-Rural Heterogeneity

Notes: The figure reports separate coefficients on sex-based grammatical gender for urban and rural respondents. Points show coefficients and horizontal bars show 95 percent confidence intervals. Specifications include the baseline controls and fixed effects where supported by the subsample. The estimates are descriptive heterogeneity results and assess whether the language association differs with institutional density and exposure to alternative gender norms.

[\[↑ back to text, Heterogeneity\]](#)

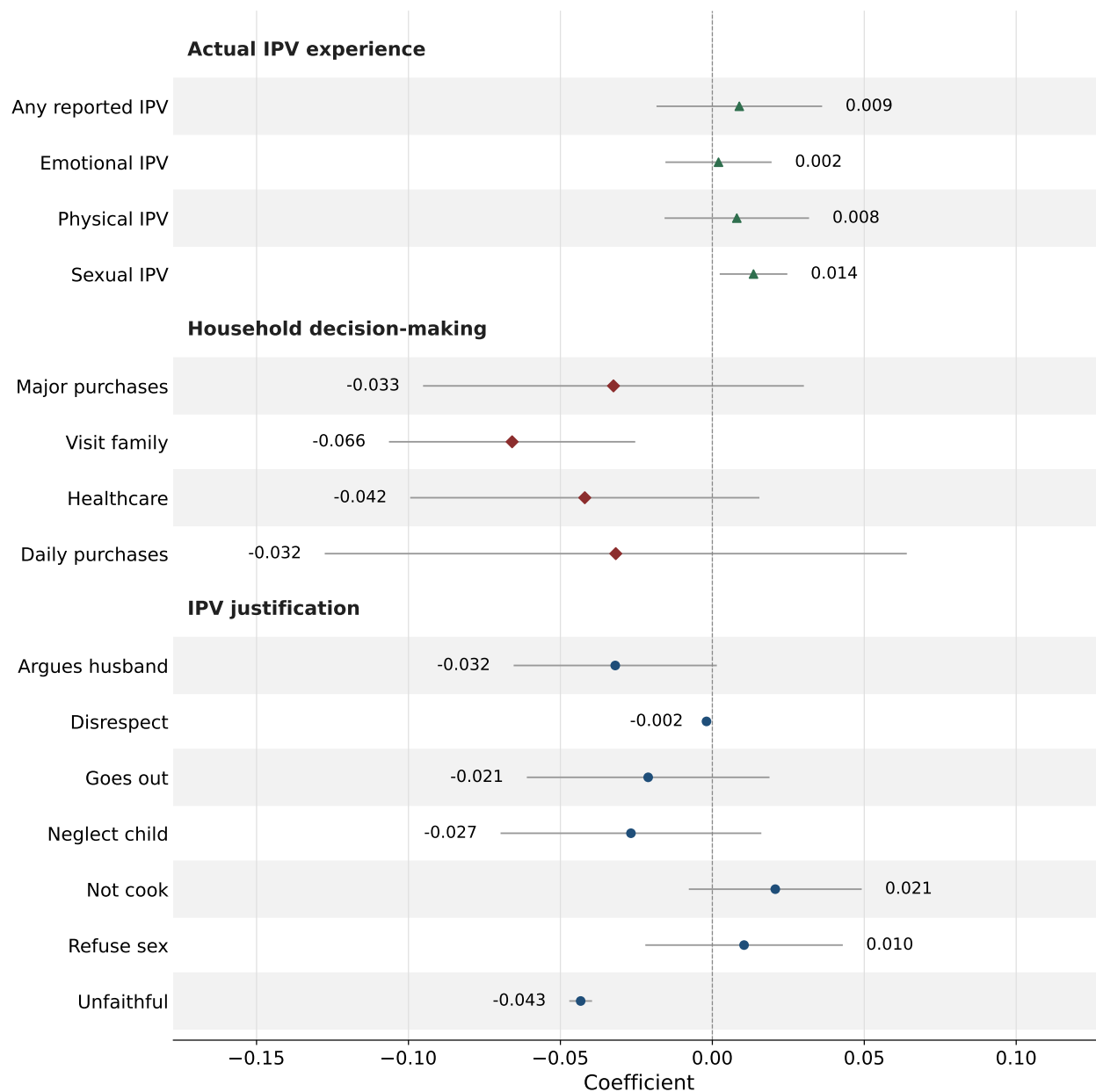


Figure A2: Future-Tense Placebo

Notes: The figure replaces sex-based grammatical gender with grammatical future tense, a linguistic feature not organized around gender. Points show coefficients and horizontal bars show 95 percent confidence intervals. Specifications follow the corresponding baseline controls and fixed effects where supported by the outcome sample. The placebo assesses whether the main associations reflect sex-based grammatical gender rather than any salient linguistic structure.

[\[↑ back to text, Robustness checks\]](#)

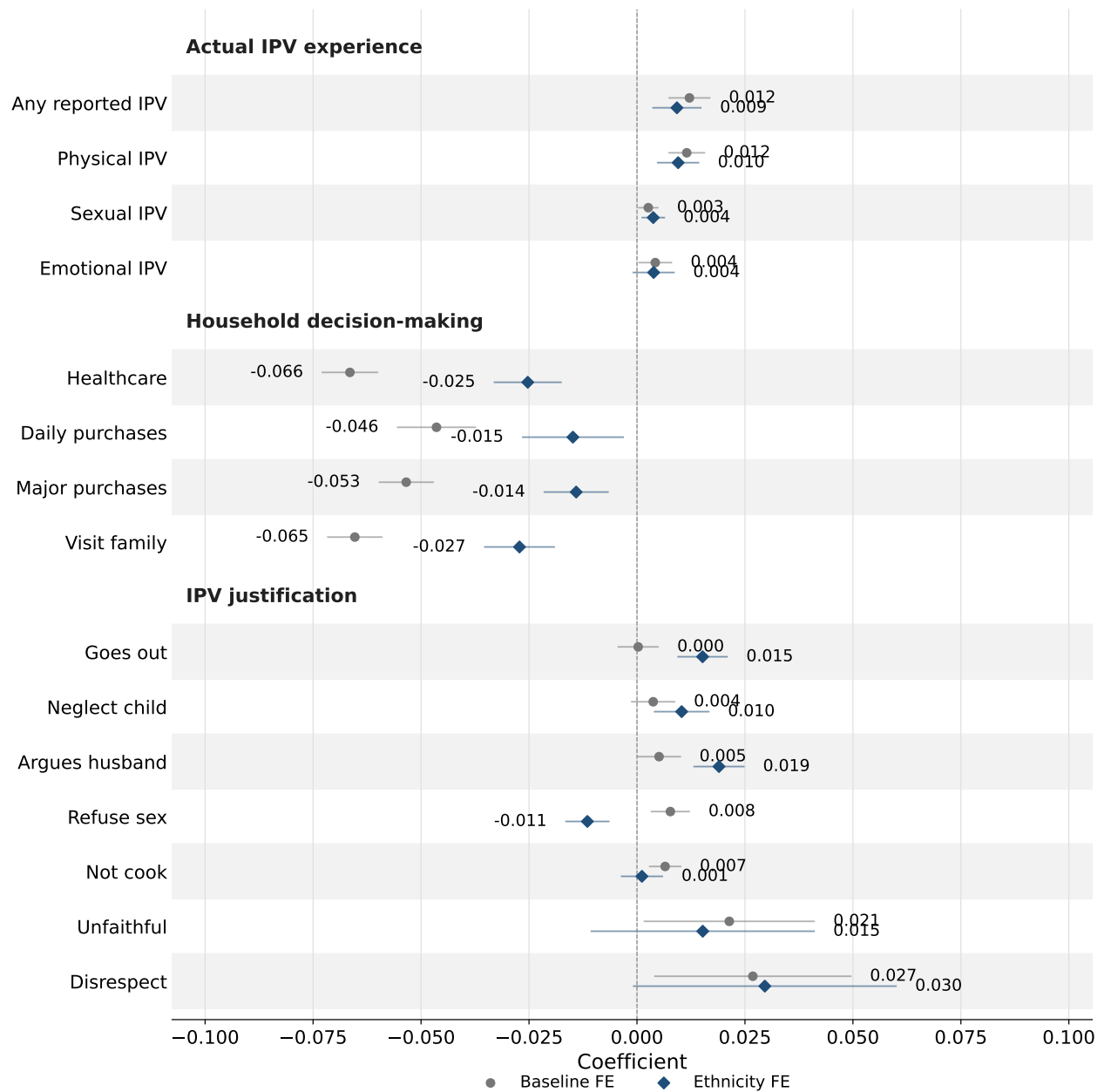


Figure A3: Ethnicity Fixed-Effects Robustness

Notes: The figure compares baseline estimates with estimates that additionally absorb ethnicity fixed effects where ethnicity is observed. Points show coefficients on sex-based grammatical gender and horizontal bars show 95 percent confidence intervals. Specifications retain the baseline controls and fixed effects apart from the added ethnicity controls. The exercise addresses concerns that ethnic-community histories may confound the language association.

[\[↑ back to text, Robustness checks\]](#)

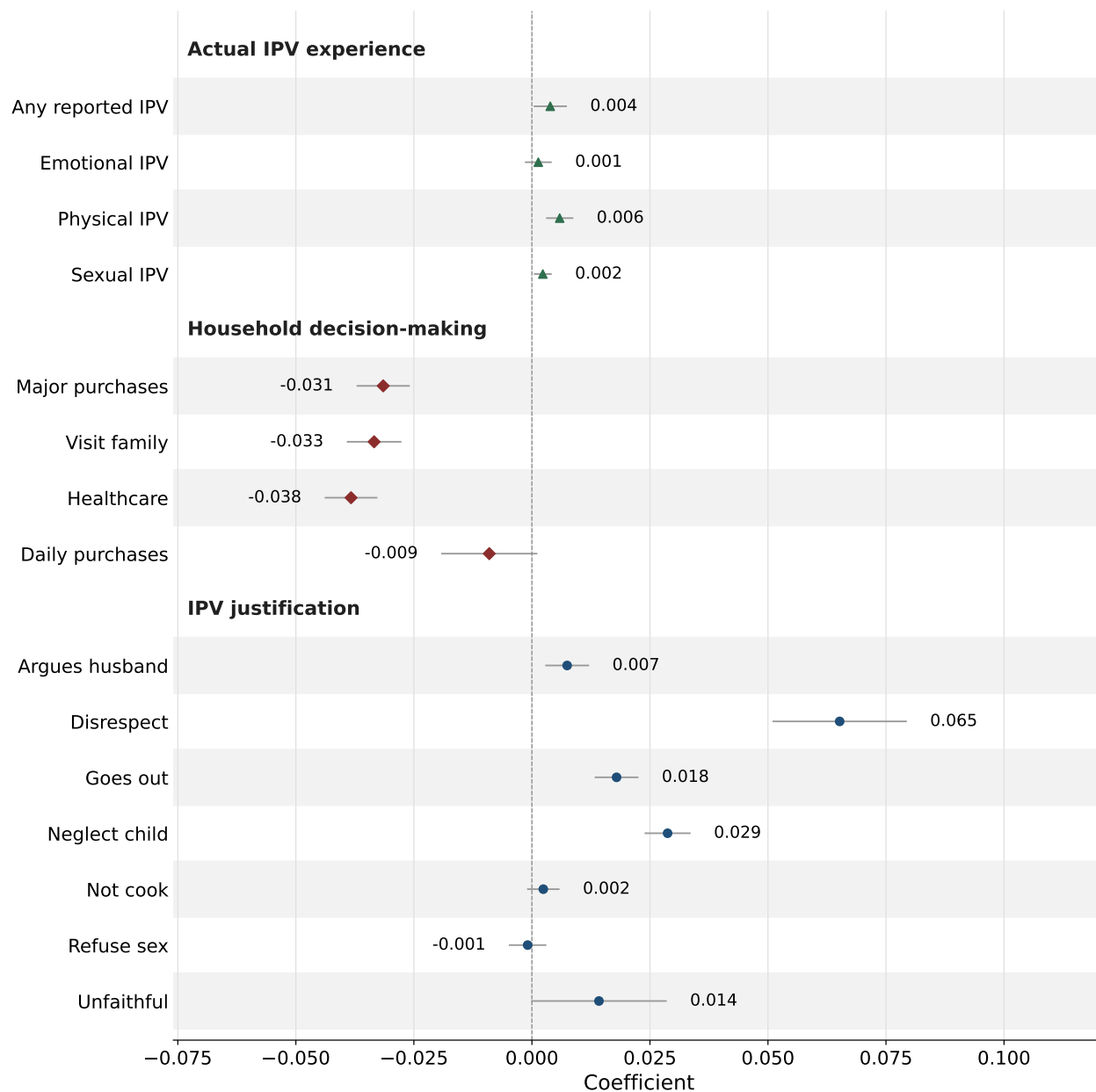


Figure A4: Province Fixed-Effects Robustness

Notes: The figure compares the baseline specification with robustness specifications that replace or augment the geographic fixed-effect structure with province fixed effects, as labeled in the figure. Points show coefficients on sex-based grammatical gender and horizontal bars show 95 percent confidence intervals. Outcomes are ordered around reported IPV experience, household decision-making, and IPV-justification channel evidence.

[\[↑ back to text, Robustness checks\]](#)

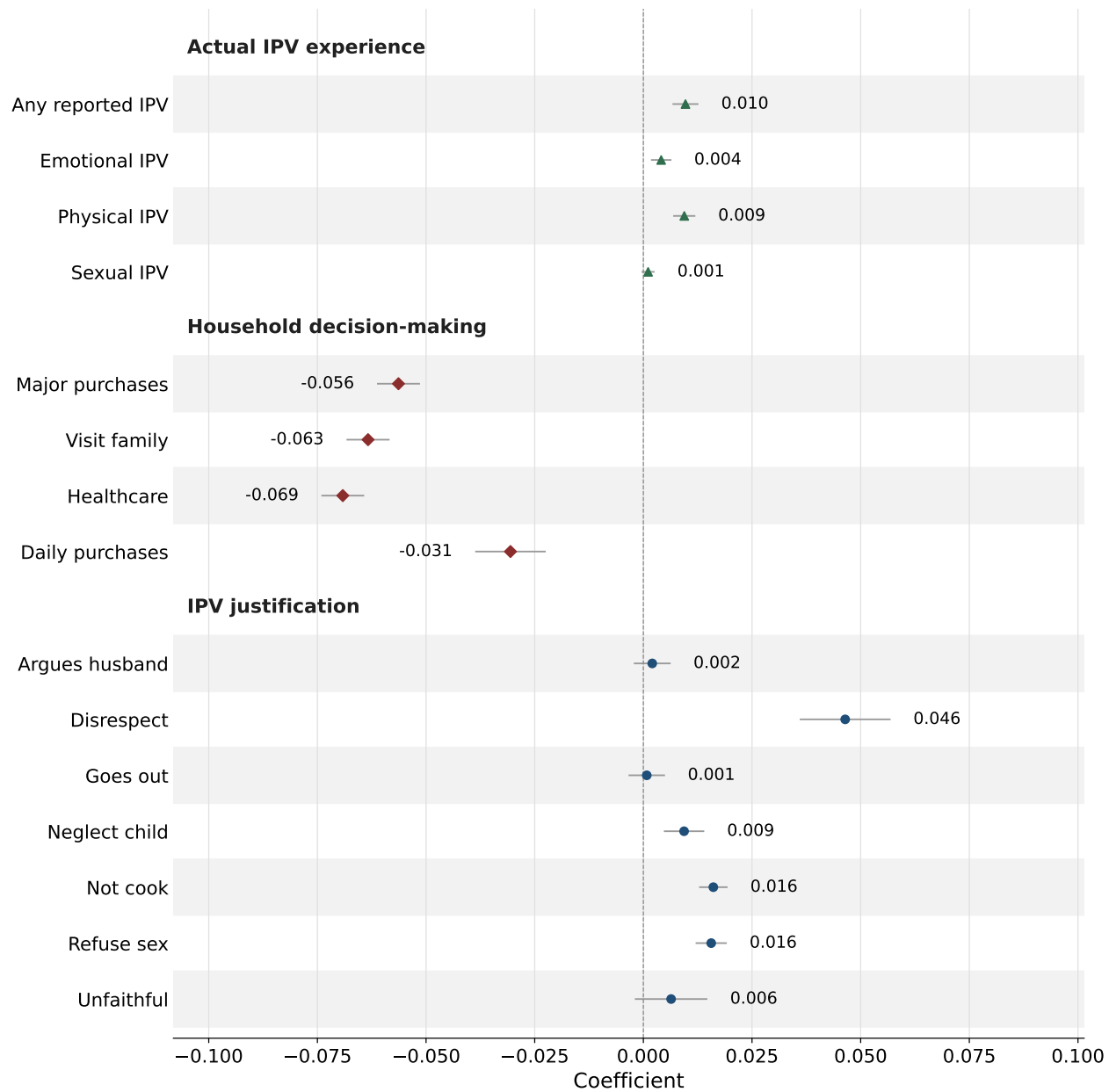


Figure A5: Country Fixed-Effects Robustness

Notes: The figure compares the baseline specification with specifications using country fixed effects, as labeled in the figure. Points show coefficients on sex-based grammatical gender and horizontal bars show 95 percent confidence intervals. Outcomes are ordered around reported IPV experience, household decision-making, and IPV-justification channel evidence.

[\[↑ back to text, Robustness checks\]](#)

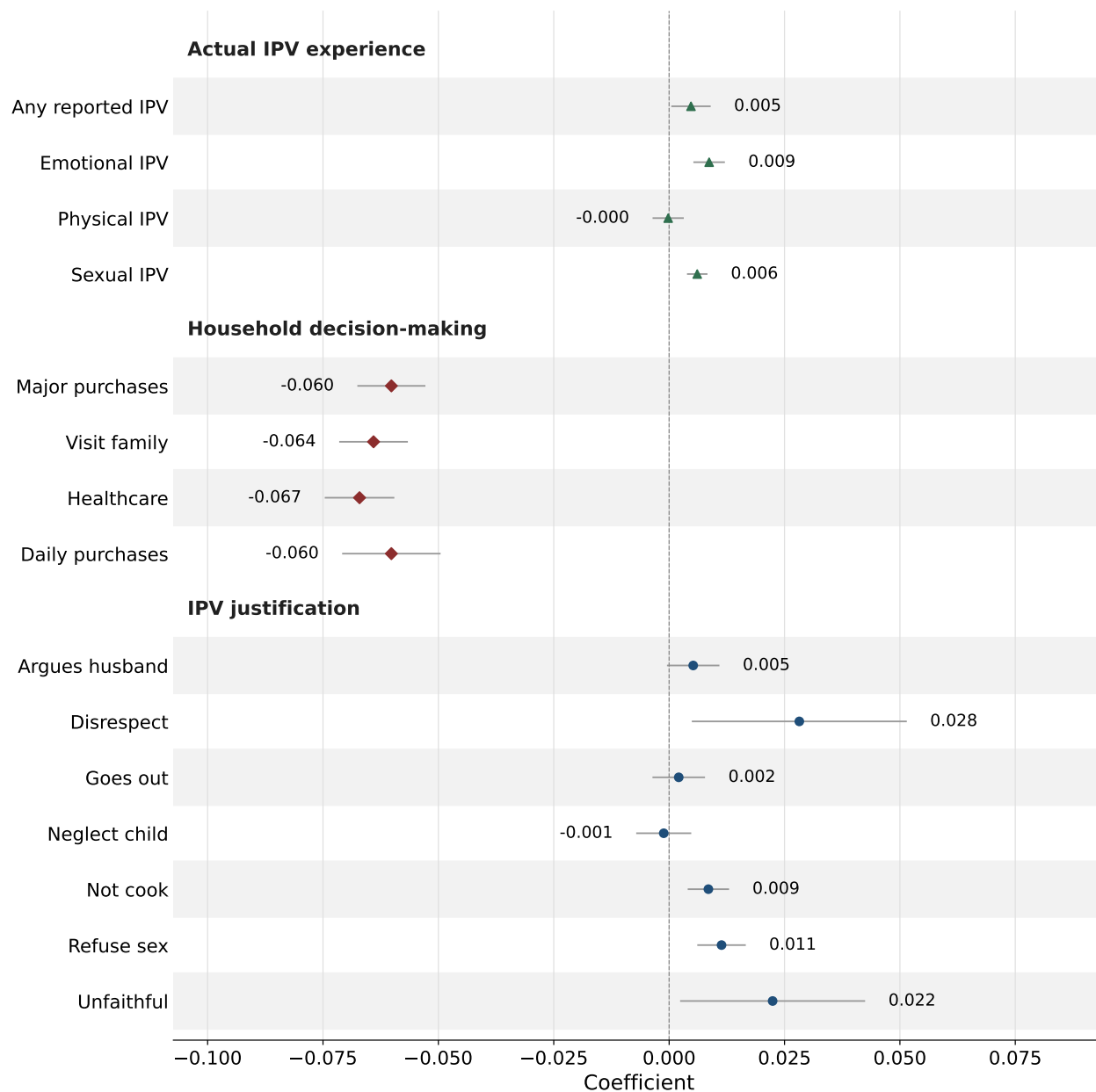


Figure A6: High-Confidence Language Classification

Notes: The figure compares estimates from the full non-conflicting language-classification sample with estimates based on stricter high-confidence assignments. Points show coefficients on sex-based grammatical gender and horizontal bars show 95 percent confidence intervals. The restriction assesses whether the main associations are sensitive to source quality, proxy assignments, or lower-confidence language classifications in the DHS-language crosswalk.

[\[↑ back to text, Robustness checks\]](#)

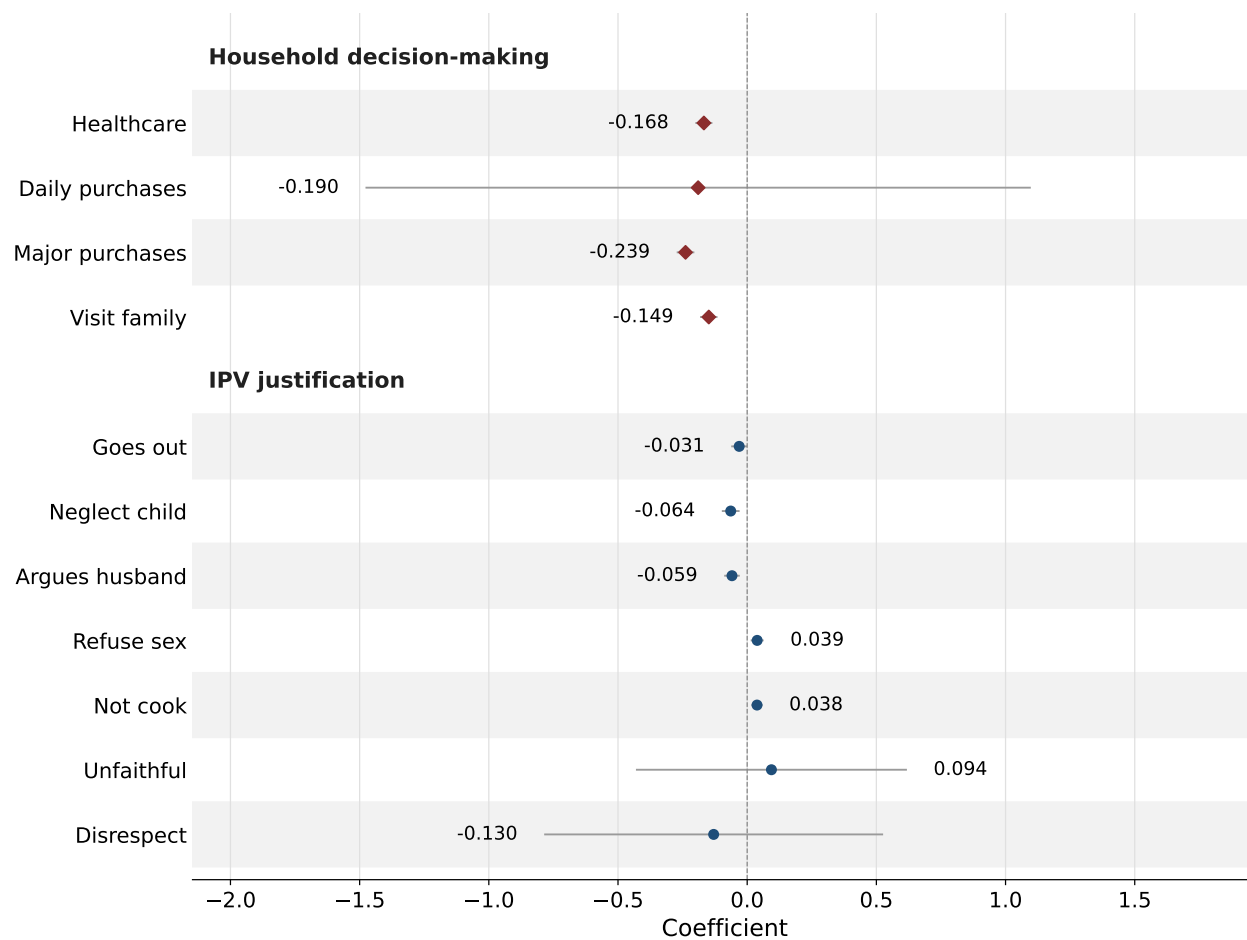


Figure A7: Component-Level IV Estimates

Notes: The figure reports component-level instrumental-variable estimates in which inherited proto-language gender structure instruments for current sex-based grammatical gender. Points show IV coefficients and horizontal bars show 95 percent confidence intervals. These estimates are sensitivity diagnostics and are not interpreted as central causal magnitudes, especially where sample size or first-stage strength is limited.

[\[↑ back to text, Robustness checks\]](#)

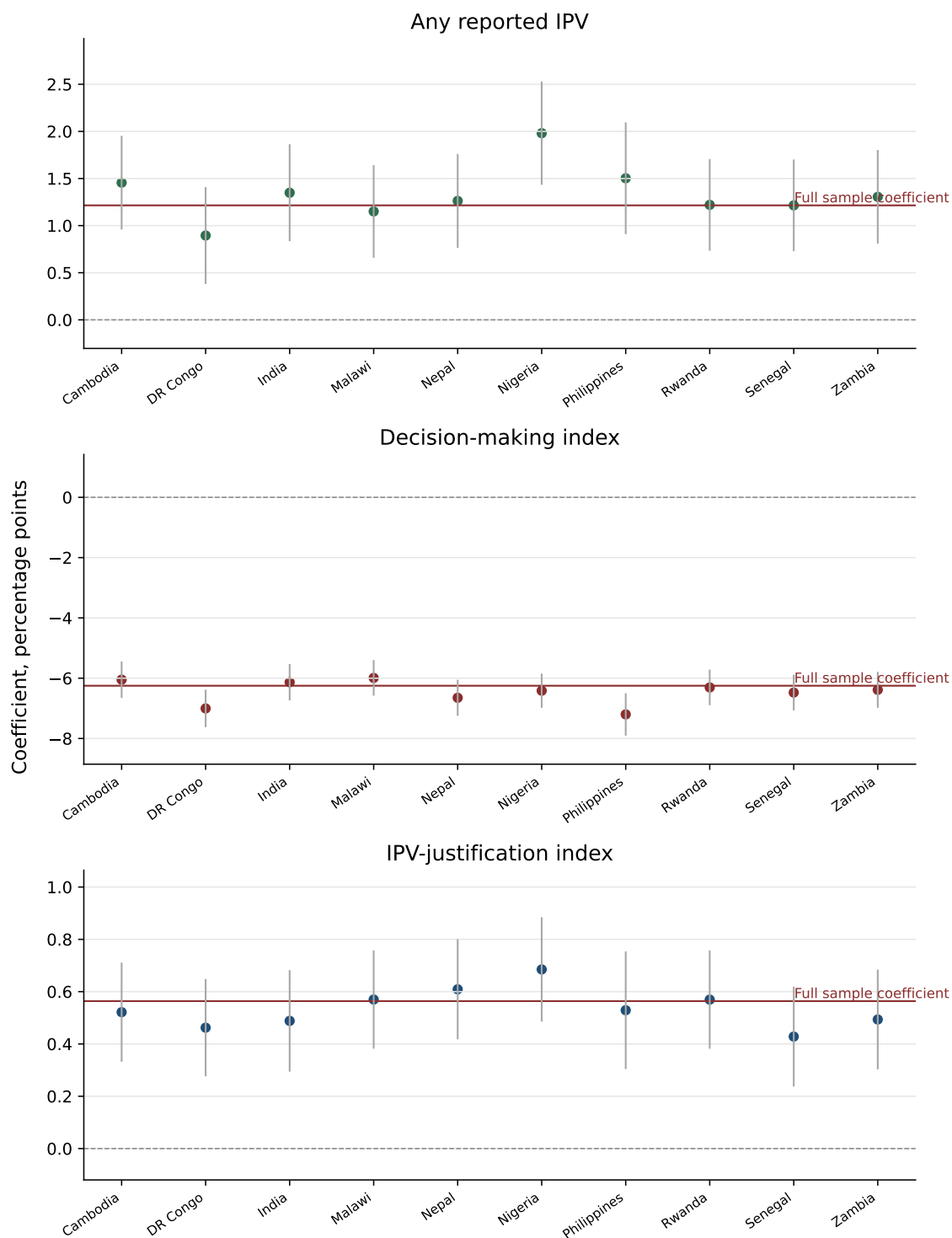


Figure A8: Leave-One-Country-Out Diagnostics

Notes: The figure reports leave-one-country-out estimates for any reported IPV, the household decision-making index, and the adjusted IPV-justification index. The ten exclusions are the largest sample-size countries in the analysis sample. Points show coefficients on sex-based grammatical gender and horizontal bars show 95 percent confidence intervals; the full-sample estimate is shown as the benchmark. Specifications follow the baseline model, and standard errors are clustered by survey cluster.

[\[↑ back to text, Robustness checks\]](#)

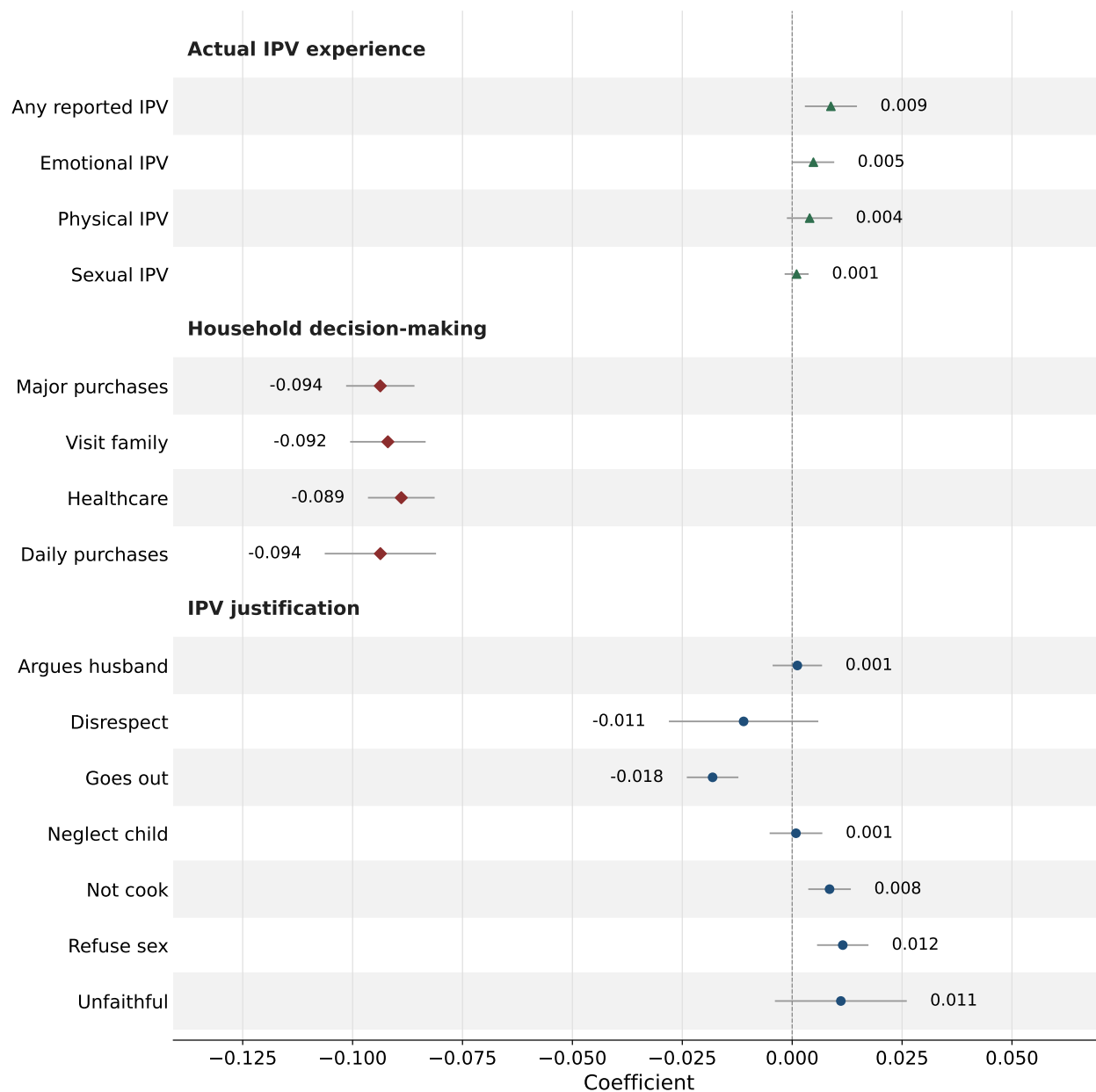


Figure A9: Formal Versus Semantic Gender Assignment

Notes: The figure reports robustness estimates using the WALS 32A distinction between formal and semantic grammatical-gender assignment. Points show estimated associations and horizontal bars show 95 percent confidence intervals. Outcomes are ordered around reported IPV experience, household decision-making, and IPV-justification channel evidence. The exercise assesses whether the results differ by the structure of gender assignment within grammatical-gender systems.

[\[↑ back to text, Robustness checks\]](#)

Appendix B. Language Extraction and Classification Process

B.1 Construction of the DHS Language Universe

The language-classification workflow began from the DHS language fields available in the harmonised analysis data. Since language information is not recorded identically across DHS countries and rounds, we first created a respondent-level primary-language label. Native or mother-tongue language was used where available; otherwise, the process used interview language and then questionnaire language. The resulting labels were collapsed to unique country-language cells so that classification occurred at the level at which linguistic identity could be checked.

We then cleaned the DHS labels before assigning typological characteristics. Cleaning involved lowercasing and trimming labels, normalising spacing and punctuation, repairing common encoding problems, harmonising obvious spelling variants, and preserving the original DHS label for traceability. Labels such as “other,” “other language,” numeric language categories, and broad residual country-specific categories were not treated as assignable languages. This step was central to data quality: the team chose not to force a grammatical-gender value onto labels that did not identify a determinate language.

Each valid label was then mapped to a canonical language identity. The matching process used country context, language aliases, DHS survey context, and linguistic identity databases. When the DHS label corresponded directly to an established language name, the assignment scope was recorded as an exact-language match. When a label was an accepted alternate spelling or local name for the same language, it was treated as an alias. When a label referred to a macro-language, language group, or broad variety name, it was handled conservatively and assigned only when the underlying source evidence was sufficiently strong and the case could be flagged as lower confidence.

B.2 Source Hierarchy and Classification Rules

The primary outcome of the classification process was the binary sex-based grammatical-gender variable. The team coded a language as sex-based if it has a grammatical gender system organised around masculine/feminine, male/female, or closely related sex-based categories. The team coded a language as non-sex-based if it has no grammatical gender or has a non-sex-based noun-class system. This includes languages with productive noun classes that are grammatical but not organised primarily around sex-based categories.

The source hierarchy was as follows. First, the preferred source was direct WALS evidence, especially WALS 31A, supported by related WALS gender-system features where available (Corbett, 2013a,b,c). WALS 30A records the number of genders; WALS 31A identifies sex-based versus non-sex-based gender systems; WALS also records whether gender assignment follows semantic, formal, or mixed criteria. Second, if WALS did not directly cover the DHS language label, the team searched Grambank, Glottolog, APiCS, and published grammars or grammar sketches for the exact language or lect (Michaelis et al., 2013; Skirgård et al., 2023; Haynie et al., 2023). Third, for languages without a single direct typological entry, the team combined identity evidence with grammar descriptions, peer-reviewed articles, theses, and other scholarly descriptions. Fourth, in limited cases, country context and closely related language evidence were used when the DHS label was not fully exact but the linguistic interpretation was strong. These cases were retained in the full coverage measure and explicitly separated from the strict high-confidence sample.

This hierarchy ensured that the most externally validated sources drove the classification. WALS is a foundational typological database constructed from descriptive materials by a large team of specialists. Grambank provides a newer, large-scale typological database with extensive coding of grammatical structures across more than 2,400 languages. APiCS provides expert-coded structural information for pidgin, creole, and mixed languages and is designed for comparison with WALS. Glottolog was used primarily for language identity, naming, and classification, while grammatical claims were based on typological databases or descriptive linguistic sources.

B.3 Human Verification and Data Quality Controls

The final language dataset was built to be reproducible and auditable. Each assigned country-language row records the canonical language, assignment scope, assignment status, confidence level, source type, citation or URL, persistent source identifier where available, and notes explaining the classification. The research team reviewed the source trail for each assigned language and checked that the final value followed from the evidence. Direct WALS matches were treated as high-confidence unless the DHS label itself was ambiguous. Exact-language grammar or Grambank evidence was also treated as high-confidence when the language identity and grammatical claim were both clear. Lower confidence was reserved for cases where the source was credible but the DHS label was broader than the exact language entry, or where the evidence came from closely related lects.⁷

⁷The team used comprehensive academic search tools, including Consensus and Semantic Scholar, to help locate potentially relevant linguistic and empirical sources. These tools supported literature discovery; final language identities, source interpretation, confidence tiers, and grammatical-gender classifications were reviewed and adjudicated by the research team.

Quality control occurred at several points. First, the team verified that all high-frequency valid language candidates were researched rather than silently left pending. Second, the team separated unassignable DHS categories from genuine language gaps. Third, the final dataset included duplicate-key checks to ensure that each country-language label merged cleanly into the analysis data. Fourth, coverage was computed on the actual estimation sample rather than only on the language-assignment file. This matters because the main regressions require nonmissing outcomes, controls, cluster identifiers, and fixed-effect structure.

B.4 Illustrative Classification Examples

Hindi. Hindi is a direct high-confidence WALS match. WALS records a sex-based system, and the project source file records Hindi as having masculine and feminine nouns with gender agreement. The DHS Hindi label in India maps exactly to the canonical language, so the row is classified as sex-based in the strict sample.

Siswati and Lozi. Siswati and Lozi illustrate the treatment of non-sex-based noun-class systems. The DHS labels map to exact language identities supported by WALS, Glottolog, Grambank, and grammar evidence. The relevant grammars describe Bantu-style noun-class systems organised into multiple classes rather than a masculine/feminine grammatical gender opposition. These languages are therefore classified as non-sex-based.

Waray. Waray demonstrates the value of triangulating WALS with Grambank and grammar sources. The final classification uses exact Grambank evidence for Waray, including the relevant grammatical feature and source citation, to classify the language as having sex-based grammatical marking for a small class of words. The row is therefore classified as sex-based on high-confidence exact-language evidence.

Gujarati. Gujarati illustrates an exact-language grammar-based classification. The DHS label maps cleanly to Gujarati, and grammar evidence describes a three-gender system with masculine, feminine, and neuter distinctions and agreement patterns. Because the classification rests on exact-language grammar evidence rather than a direct WALS 31A entry in the project extract, it is retained as a medium-confidence sex-based assignment.

B.5 Coverage in the Main Estimation Sample

The coverage figures below are computed for the largest main-outcome estimation sample in the language-extraction audit. This sample contains 1,531,703 observations after applying the normative-channel index specification and the `reghdfe` estimation sample. The broader main-analysis universe contains 1,536,518 observations, but [Table B1](#) and [Table B2](#) use the language-extraction audit sample as the denominator.

Table B1: Language-classification coverage by source method

Source method	Observations	Share of audit sample
Direct WALS assignment	1,196,869	78.1%
Exact Grambank assignment	111,457	7.3%
External exact grammar/database assignment	107,661	7.0%
Source-backed proxy or lower-confidence assignment	96,062	6.3%
Local project source	19,654	1.3%
Total audit estimation sample	1,531,703	100.0%

Notes: The table reports the source method used to assign sex-based grammatical-gender status to observations in the language-extraction audit sample. Shares use the audit estimation sample as the denominator.

[\[↑ back to text, Appendix B\]](#)

Table B2: Language-classification coverage by quality filter

Quality filter	Observations	Share of audit sample
Strict/high-confidence, non-proxy, non-conflicting	1,249,194	81.6%
High-confidence, non-conflicting	1,249,194	81.6%
High- or medium-confidence, non-conflicting	1,397,677	91.2%
Full non-conflicting assigned sample	1,531,703	100.0%

Notes: The table reports quality-filter coverage for the same estimation sample as the source-method coverage table. Lower-confidence assignments are retained only where the source trail supports a non-conflicting classification, with the confidence tier preserved for robustness checks.

[\[↑ back to text, Appendix B\]](#)

These coverage figures show that the main analysis is not driven by undocumented language imputations. Most observations are classified directly from WALs, and more than nine-tenths of the audit estimation sample is covered by high- or medium-confidence evidence. The remaining observations are retained only where the source trail supports a non-conflicting classification, with the confidence tier preserved for robustness checks.