Beblo, Miriam; Görges, Luise; Markowsky, Eva

Working Paper

Gender matters in language and economic behaviour

WiSo-HH Working Paper Series, No. 48

Provided in Cooperation with:
University of Hamburg, Faculty of Business, Economics and Social Sciences, WISO Research Lab

Suggested Citation: Beblo, Miriam; Görges, Luise; Markowsky, Eva (2019) : Gender matters in language and economic behaviour, WiSo-HH Working Paper Series, No. 48, Universität Hamburg, Fakultät für Wirtschafts- und Sozialwissenschaften, WiSo-Forschungslabor, Hamburg

This Version is available at:
https://hdl.handle.net/10419/260451

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

FAKULTÄT
FÜR WIRTSCHAFTS- UND
SOZIALWISSENSCHAFTEN

Gender matters in language and economic behaviour

Miriam Beblo
Luise Görges
Eva Markowsky

WiSo-HH Working Paper Series
Working Paper No. 48
April 2019

WiSo-HH Working Paper Series
Working Paper No. 48
April 2019

Gender matters in language and economic behaviour

Miriam Beblo, Universität Hamburg
Luise Görges, Universität Hamburg
Eva Markowsky, Universität Hamburg

ISSN 2196-8128

Font used: „TheSans UHH“ / LucasFonts

Kontakt:
WiSo-Forschungslabor
Von-Melle-Park 5
20146 Hamburg
E-Mail: experiments@wiso.uni-hamburg.de
Web: http://www.wiso.uni-hamburg.de/forschung/forschungslabor/home/

Gender matters in language and economic behaviour

Miriam Beblo, Luise Görges, Eva Markowsky*
Universität Hamburg

April 17, 2019

Abstract

We study the role of gendered language for gender differences in economic behaviour. Previous studies posit that speakers of gendered languages make more gender-stereotypical choices because speaking requires them to reference gender more frequently compared to speakers of neutral languages. A recent literature in economics has attempted to isolate this direct effect of language on behaviour from those of institutions and other cultural influences. Against this background, we incorporate the linguistic theory of "thinking for speaking" into an identity economics model of labour supply and show how speaking a gendered language may affect gender differences in economic outcomes. At the same time, the model highlights pitfalls in estimating this effect empirically. A systematic literature review and our own empirical exercises illustrate the severity of the problem, in particular with respect to the epidemiological approach, i.e. studying migrants. While several studies adopted this empirical strategy with the intention to combat endogeneity issues in measuring a causal effect of language on economic behaviour, our model and analyses suggest that it may in fact make matters worse. Using a European data set, we show that migrant behaviour seems hardly related to speaking a gendered language, regardless of the empirical specification, and argue that this finding is driven by self-selection into migration for first generation immigrants and intergenerational transmission of social norms and values for the second generation. Overall, we conclude that the epidemiological approach is not suited to solve the issue of co-evolvement of culture and language, and, moreover, does not seem to solve the issue of self- and parental selection.

*Valuable comments by Gerd Mühlheußer, Grischa Perino and Arne Pieters are gratefully acknowledged.

1 Introduction

Despite vast improvements in gender equality over the past several decades, women are still less active in the labour market and spend fewer hours in paid work than men (UNDP, 2015: 108). Although this is true in most countries of the world, the extent to which women integrate into the labour market varies widely. For instance, the 2017 employment rate of women ranged from only 29% in Turkey to almost 77% in Iceland (OECD, 2019). In the past decade, economists have begun to investigate the role of social norms and culture in shaping these differences (e.g., Akerlof & Kranton, 2000; Duflo, 2012; Fortin, 2005; Fernández & Fogli, 2009). Even more recently, research has focused on the role of language (e.g., Mavisakalyan, 2015; Gay *et al.*, 2017) and the question to what extent cross-linguistic differences in gender marking could potentially help explain variations in individual economic behaviour. Two channels are plausible: The first runs from grammatical structure through cognition to speakers' choices. The second runs from institutions and culture to individual behaviour. Assuming that culture and institutions (including language) co-evolve, the gender marking of a language is indicative of the degree of gender inequality in a society. While not mutually exclusive, the second channel presents a challenge for empiricists trying to provide evidence for the first: a direct, causal effect of language on behaviour.[1]

The present paper seeks to further develop our understanding of the role of gendered language for gender differences in economic outcomes in three respects: i) We develop an economic model for understanding the effects of language on labour supply; ii) We review empirical studies in the economics literature that investigate the link between gendered language and gender gaps in various economic outcomes, and identify challenges for establishing causality; and iii) We present an empirical analysis of European labour markets that confirms the challenges we have outlined in determining causal relationships.

We begin by incorporating insights from (psycho-) linguistics into an identity economics model of labour supply. In the model, individuals may distort their labour supply to match the social prescription of an "ideal" level of participation for their gender, which is given by a gender norm. As gendered language makes gender norms more salient, the distortion will be stronger for individuals who speak languages that mandate reference to gender more frequently. We use the model to show how an effect of language on behaviour can be incorporated into economic analyses and, moreover, to illuminate the challenges we might face in identifying such an effect empirically. If gender norms and gendered language are correlated across speaking communities, gender differences

[1] The idea that language may influence economic choices has received recent attention by economists. In the most renowned paper, Chen (2013) provides empirical evidence that speakers of "futureless" languages exhibit more future-oriented behaviours (saving money, not smoking, practicing safer sex, maintaining a healthy weight). He argues that they perceive present and future as more closely associated, since their languages' grammar do not (consistently) require separating between the two time horizons. His argument can be formalized in a theoretical framework similar to the one we present in this paper, as the mechanisms he proposes also go from language through cognition to behaviour.

in labour supply might be larger among speakers of gendered languages both because gender norms are more salient but also because the norms themselves prescribe larger differences in the "ideal" behaviour for men and women.

Our review of empirical papers in the economics literature confirms that the two channels are not easily discriminated, neither conceptually nor empirically. Even though most authors are interested in the direct, cognitive effect of language on behaviour, the possibility that effects are driven by culture and, in some cases, institutions, cannot be dismissed. The problem is most apparent in cross-country studies that compare economic outcomes at the country level (e.g., the gender pay gap) or at the individual level (e.g., labour force participation), and find correlations between these outcomes and the gender marking in the countries' dominant languages. Seeking a remedy, researchers have turned to the epidemiological approach (Fernández & Fogli, 2009) and studied the behaviour of immigrants, for whom the institutional context at the time of making decisions of interest no longer coincides with the institutional context in which their language evolved. In the context of gendered language, however, this approach creates new problems of endogeneity (and ambiguity, as shows our model) that we find warrant further cautious investigation.

In the empirical section of the paper, we demonstrate these identification challenges in an application of the epidemiological approach to the labour market behaviour of immigrants in Europe. To this end, we construct a dataset that is more comprehensive than the ones used in previous studies, and allows us to compare different linguistic indicators that distinguish between gender-intensive and gender-neutral languages. We use several waves of the European Social Survey (ESS) and merge them with linguistic data regarding the respondents' languages. We obtain information about four features of grammatical gender for each language from the World Atlas of Language Structures (WALS) (Dryer & Haspelmath, 2013), and complement this data with information from further linguistic sources for languages missing from the WALS database. Conclusions regarding the extent to which language plays a role in shaping behaviour vary, depending on the linguistic indicator used, how the sample is constructed (first- vs. second-generation immigrants), and how language is assigned (most commonly spoken at home vs. dominant language in country of origin). Generally, our results do not confirm the leading hypothesis in the literature, which is that behaviour is more consistent with gender stereotypes for speakers of a gendered language. Instead, our application of the epidemiological approach suggests a weak correlation between behaviour and speaking a gendered language. While for men we find a negligible effect regardless of whether we look at first or second generation immigrants or whether we assign language by the dominant language in the country of origin or by the language respondents speak at home, women's labour force participation is significantly related to the dominant language for some definitions of gender marking only. If anything, the estimated impact of a gendered language on labour force participation is positive; i.e., female respondents whose parents emigrated from countries with gender marking in their dominant language

tend to work more, which means that their behaviour is *less* stereotypical. We argue that these results are consistent with the notion that language reflects cultural traits, and that men and women who select into migration (first generation) are more likely to reject the culture in their country of origin, and may transmit this attitude to their children (second generation). While the epidemiological approach has its advantages in separating the effects of culture from those of institutions, it is not well-suited to tease out the cognitive effects of language from those of culture, and neither does it seem to overcome self-selection and parental selection issues.

Our paper is most closely related to a recent paper by Mavisakalyan & Weber (2017) who provide a comprehensive overview of the economics literature on linguistic structures as determinants of economic behaviour. Mavisakalyan & Weber (2017) discuss several channels through which cross-linguistic differences could influence decision-making in an economic framework, many of which are also important in our model. Our model goes beyond their framework by formalising the effects of language on a very specific decision, the individual supply of labour. It thus provides a tool for integrating language into economic analyses that can easily be applied to other contexts. Mavisakalyan & Weber (2017) also provide a broader review of the literature than we do, including summaries of the empirical evidence on economic outcomes that relate to linguistic differences regarding not only gender but also tense, pronoun use, and moods. While they discuss issues of identification extensively, the goal of our paper is to focus the discussion on the question whether the epidemiological approach, which is gaining popularity in economic studies of language, presents a solution to the challenge of isolating a causal effect on economic behaviour. From our model we develop a theoretical argument against this idea and review the empirical literature accordingly. Finally, we complete the picture by presenting the results of our own empirical analyses with European data.

The paper is organized as follows: The theory Section **??** lays out insights from linguistics and psycho-linguistics, from which we derive an identity economics argument how gendered language can affect behaviour. Section 3 presents an overview of economic empirical studies on the subject and highlights inconsistencies in the results, as well as challenges to their interpretation. In Section 4, we present our empirical application, including data description and empirical approach. The results in Section 5 support the conclusions we derived from the theoretical framework and the systematic literature review. Finally, Section 6 offers a summary and concluding remarks.

2 Theory

In this section we develop an identity economics model that builds on the (psycho-)linguistic insights we describe in the first subsection.

2.1 The linguistics background

Gender is embedded into the linguistic structure of a language in different ways, as both a grammatical category and a linguistic representation of sex. Generally, linguists distinguish between formal grammatical gender and semantic natural gender (MacKay, 1999).

Stahlberg *et al.* (2007) differentiate among three types of languages, based on their linguistic representation of biological sex: grammatical gender languages, natural gender languages, and genderless languages. In grammatical gender languages—e.g., Spanish, French, and German—sex is coded as a grammatical category. Every noun is assigned either female, male, or (in some languages) neutral gender, and articles, adjectives, pronouns, and, in some cases, even verbs must agree with the gender of the noun to which they refer (Boroditsky *et al.*, 2003, among others). For the majority of personal nouns, grammatical gender corresponds with the sex of the person referred to. Thus, a distinction such as "waiter/waitress," which is the exception in the English language, exists for nearly every personal noun in grammatical gender languages. Consequently, these languages involve frequent reference to gender when verbalizing statements about human beings (Stahlberg *et al.*, 2007). Such references are less frequent in natural gender languages, because they include no explicit grammatical marking of sex, and gender is used mostly semantically (Konishi, 1993). Examples are English and many Scandinavian languages. With a few exceptions, as in the aforementioned example of "waiter/waitress," personal nouns can refer to both women and men. Only gendered pronouns require referencing sex (he, she) (Stahlberg *et al.*, 2007). Genderless languages—e.g., Finnish and Hungarian—feature grammatical gender neither in their noun system nor in pronouns. Gender is expressed through lexical means only (e.g., in the words for father, mother, sister, brother, and the like) and references to sex are therefore extremely rare (Stahlberg *et al.*, 2007).[2] Some researchers point out that a strict division between these categories is not possible since natural gender languages can feature aspects of grammatical gender languages and vice versa (Konishi, 1993; MacKay, 1999).

The idea that these language structures influence the cognitive process dates back to the end of the 19th century, although it gained the most attention through the infamous Sapir-Whorf hypothesis of the 1950s (Casasanto, 2016: 158). According to this hypothesis, also known as Linguistic Relativity or Linguistic Determinism, a person's native language determines her perception, thought, and action (Whorf, 1956). While most researchers have rejected the hypothesis in its strongest, original formulation, a modified version put forward by the psychologist and linguist Dan Slobin (1987, 1996) has found much empirical support. Relative to the original hypothesis, Slobin and other authors suggest a more moderate effect of language on cognition, in the sense that "language differentially favors some thought processes over others" (Hunt & Agnoli, 1991: 377-378). According to Slobin (1996), language requires "thinking for speaking"; i.e., the cognitive process of verbalizing experiences. Because of their different grammatical structures, languages vary in the degree to which they draw the speaker's attention to certain aspects in this process. (Slobin, 1996, 71) illustrates this idea by means of the statement "The man is sick." Some languages mandate that the speaker indicates, using

[2] According to Stahlberg *et al.* (2007), entire language families fall into these three types: (1) Grammatical gender languages: Slavic, Germanic, Romance, Indo-Aryan, and Semitic languages; (2) Natural gender languages: English and the Scandinavian languages; and (3) Genderless languages: Uralic, Turkic, Iranian, Sinitic, and Bantu languages.

grammatical terms, whether the sickness is temporary or chronic, while speakers of other languages would have to specify whether the man was moving or at rest, and so on. He argues that such obligatory reference to certain categories will affect the representation of these categories in the speakers' mind – at least in the process of generating or receiving verbal messages. He refers to gender as another example of semantic features that require obligatory reference in some languages but not in others, indicating that speaking a grammatical gender language draws the attention to gender while "thinking for speaking" (Slobin, 2003: 160).

Such a powerful influence on the thought process would not only stem from languages' grammatical idiosyncrasies, forcing speakers to reference certain categories yet ignore others, but from languages' ubiquity in everyday life. Humans are "almost constantly involved in preparing, producing, and interpreting verbal messages" (Slobin, 2003: 158). Moreover, as Casasanto *et al.* (2004) argue, people might form cognitive habits while "thinking for speaking" which means that these habits become a routine even at times when they are not preparing verbal utterances. Although the authors study mental representation of time, the argument extends to grammatical gender. If a language forces speakers to sort objects into certain categories according to its grammatical rules, it essentially defines which categories deserve the speakers' attention. With respect to grammatical gender languages, this would imply that speakers experience the social categories of female and male as much more relevant than speakers of gender-neutral languages.

2.2 A psycho-linguistic identity economics model

If the frequent categorization by gender mandated in grammatical or natural gender languages increases the salience of gender as a social category, it may conceivably affect economic decision-making. Consequentially, speakers of grammatical gender languages may exhibit more gender-stereotypical decision making[3]. We build on Akerlof & Kranton's (2000) identity theory to formalize this argument. The authors' original model integrates social norms as a key determinant of individual behaviour into a standard utility maximization framework. Social norms become powerful because individuals internalize them; i.e., they become an integral aspect of individual preferences. Specifically, the identity utility of an individual—a member of a social category—is directly

[3]The notion that gender-specific language leads to lower levels of gender equality is somewhat at odds with feminist language critique, that frequently calls for *more* gender-differentiating in language to enhance women's visibility and thus gender equality, e.g. in job titles. Even though masculine generics are far more common in grammatical gender languages, they also exist in natural gender languages (e.g. expressions such as "Man is mortal" in English) and even in genderless languages (e.g. the Finnish term for "civil servant": "virkamies", with the ending –mies literally translating to "man" or "male" (Tainio, 2006)). In the spirit of linguistic relativity, genderless and natural gender languages should call less attention to gender and sex even if they are not completely symmetrical in their treatment of men and women. Any empirical measure of cognitive or behavioural differences between speakers of gender-neutral and gendered languages should therefore be considered a lower bound. From a policy perspective, the question of what works better (or what is more feasible) – making language more gender neutral or more gender symmetrical– is ultimately an empirical one. For a comprehensive discussion of feminist language critique, see (Stahlberg *et al.*, 2007, 170-174).

affected by her ability to match the behavioural prescriptions for that category.[4] The authors illustrate their theory against the background of several examples, with gender inequality in the labour market being the most relevant in the context of the present paper. Therein, the relevant social categories are "man" and "woman" and the social norms attached to these categories influence identity utility by prescribing behaviour that is considered appropriate in the labour market.

Notably, Akerlof & Kranton (2000) emphasize that "when an individual's identity is associated with multiple social categories, the 'situation' could determine, for example, which categories are most salient" (731). Following this reasoning and the psycho-linguistic perspective, we hypothesize that speaking a gender-intensive language, which forces the speaker to pay more attention to the social category of gender, increases the salience of this category. Assuming that individuals' (economic) choices are generally influenced by prevalent gender norms and behavioural prescriptions, speakers of gender-intensive languages should be more likely to match behavioural prescriptions for the category of gender and thus more likely to make gender-stereotypical decisions.

2.3 The salience of gendered language

We modify Akerlof & Kranton's (2002) schooling model (AK) to formalize this notion. While in the AK model the salience of identity is given exogenously, we allow it to depend on language. We assume that an individual's utility directly depends on her efforts in the labour market and the monetary returns to it.[5] Denote by e_i the individual's effort in the labour market, measured in hours worked and thus observable to the researcher as labour supply.[6] $c(e_i)$ is the pecuniary costs of effort, including forgone leisure or home production. A standard utility function would thus describe utility as a function of income and effort only: $U_i(we_i, c(e_i))$, where w is the wage rate per unit of effort. If we assume men and women in a society to be identical, i.e., to differ neither in terms of their returns to effort, effort costs, nor preferences, all agents in the economy will supply the same amount of effort to the labour market. This framework can be augmented by integrating social categories and the corresponding behavioural prescriptions such that the standard utility function described above becomes: $U_i(we_i, c(e_i), I_i)$, where $I_i = I_i(e_i; e^G)$ is the identity component of the utility function. Note that the relevant behavioural prescriptions for specific social categories here are assumed to be the "ideal" effort levels for individual i of gender G (where $G = f, m$), which we denote by e^G. As in the AK model, the standard utility and the identity component are combined using

[4] An agent's identity utility can also be affected by externalities, i.e., by other people's behaviour. This creates additional social pressure to comply with prevalent societal norms. We will discuss this aspect later in this section.

[5] For simplicity, we follow the original set-up in the AK paper as closely as possible, except that we change the interpretation of the parameters to apply to the context of labour market choices of men and women, rather than schooling choices of "jocks", "nerds", and "burn-outs".

[6] We are aware that effort supply can vary per unit of time, an argument put forward by Becker (1985). However, for the sake of following AK's model set up as closely as possible, we keep the variable effort rather than using time supplied to the market, and assume that effort can be measured empirically as hours worked.

a weighting function. For convenience, we employ the simple functional form used by AK, as it suffices to illustrate the main insight, and express utility for individual i as follows:

$$U_i^G = (1-p)\left[we_i - \frac{1}{2}e_i^2\right] + p\left[-\frac{1}{2}(e_i - e^G)^2\right] \tag{1}$$

The parameter p ($0 \leq p \leq 1$) denotes the weight that is placed on the conformity with one's social category. Individuals experience an identity penalty whenever their effort supply deviates from the socially prescribed, "ideal" effort level for their gender, e^G. The two polar cases of p are worth examining. When $p = 0$, social categories, and thus identity concerns, are irrelevant to the individual's utility. To maximize utility, subjects simply choose the effort level that equalizes marginal benefit to marginal cost of effort, so $e_i^* = w$. When $p = 1$, the individual will choose an effort level that matches the social prescription exactly, $e_i^* = e^G$. Assuming for now that the behavioural prescriptions for the social categories "man" and "woman" are exogenously given, and that $e^f < e^m$, men would supply more effort to the labour market than women. In fact, for $e^f < w \leq e^m$, this conclusion carries over for any $p > 0$, as the optimal effort level is given by $e_i^* = (1-p)w + pe^G$.

The linguistic theories on "thinking for speaking" suggest that p may be a function of the decision-maker's language, $p(l)$. We assume that it increases in the frequency with which it forces speakers to reference the category of gender. For given gender norms, i.e. e^f and e^m are the same across countries, this implies that effort choices are more in line with gender norms for speakers of gendered languages than for speakers of gender-neutral languages. To see this, imagine that l were continuous, i.e., language's gender marking may fall anywhere on the spectrum between completely neutral and fully gendered. Then, the first derivative of the optimal effort level e_i* with respect to l can be examined to see how higher gender marking affects effort supply:

$$\frac{\partial e_i^*}{\partial l} = (e^G - w)\frac{\partial p}{\partial l}. \tag{2}$$

Because we assumed that $e^f < w \leq e^m$, and that the perceptiveness to gender norms, p, increases in l, this means that the difference in optimal effort supply between genders increases as l increases.

2.4 Institutions and societal gender norms

So far, our model predicts that cross-country variation in the gender gaps in labour supply emerges from variation in the grammatical gender feature of languages while assuming that gender norms themselves are invariant across countries. When we relax this assumption and allow heterogeneity in norms, cross-country variation in the gender gap may still result from the mechanism described by the model, i.e., gendered

languages increases the salience of gender categories, as long as the gender norms in a country are orthogonal to the presence of a gendered language. This premise, however, appears hard to justify. It seems much more plausible that language, institutions, and culture, co-evolve. In the linguistics literature, several studies emphasize that culture is a problematic confounder when attempting to isolate the causal effect of language on cognition (e.g., Beit-Hallahmi *et al.*, 1974; Roberts & Winters, 2013; Beller *et al.*, 2015).[7] Consider low female labour force participation and language gender marking as an example: If a society, historically, strongly relied on a gendered division of labour, this may have shaped both its language and institutions. The gender marking would then simply reflect the (historical) pervasiveness of gender norms in the speech community, which will also be confounded with current institutional characteristics, such as low support for female labour force participation. Recently, economists were able to provide evidence that this concern is justified: Galor *et al.* (2017) show that specific geographical conditions, determined well before industrialization, predict certain characteristics of the local languages, such as the existence of a future tense or the presence of grammatical gender. Because these conditions have also been shown to affect institutions and the way society has been organized historically, the authors conclude that language structures reflect past human experience and ancestral cultural traits.[8] Thus, at the country level, the singular impact of language on cognition and individual choices is hard to establish because p and the difference $e^m - e^f$ will be positively correlated.

We can incorporate the idea that language may be shaped by social norms into our model by allowing for the prescribed effort levels, e^G, to depend on language as well. Extending the model in this manner yields[9]:

$$U_i^G = (1 - p(l))\left[we_i - \frac{1}{2}e_i^2\right] + p(l)\left[-\frac{1}{2}(e_i - e^G(l))^2\right] \tag{3}$$

While this modification does not alter the condition for an optimal effort level, it *does* change the first derivative of the effort supply function with respect to the gender

[7] Even languages themselves may be inter-related through cultural evolution. Common ancestors or spillovers can cause neighbouring languages to share linguistic features (Roberts *et al.*, 2015)

[8] Notably, some studies in the economics literature go one step further and assume that social norms and language are linked only through historical conditions, dismissing the possibility that language may affect cognition entirely. These studies argue that grammatical features of languages, because they evolve slowly over time, are only correlated with values and social norms, but do not directly affect current institutional outcomes. For example, Tabellini (2008) exploits the presence of a certain grammatical feature, the rules of pronoun use, as an instrument to estimate the causal effect of morality as a social norm on the quality of government. Givati & Troiano (2012) use the number of gender-differentiated pronouns to instrument gender discriminatory attitudes to show the causal effect of the latter on maternity leave policy. Thus, these authors take an even stronger stance and refute the possibility that language may affect cognition, because this would mean it could, in fact, also have an influence on current policies and institutions.

[9] To keep matters tractable, we use a static model to illustrate our main point. The fact that languages and norms co-evolved in the past affects the individual decision-maker in the present only inasmuch as both social gender norms, e^G, and the weight she places on gender identity, p, are determined by l.

marking of the language. Formally, it becomes:

$$\frac{\partial e_i^*}{\partial l} = (e^G - w)\frac{\partial p}{\partial l} + p(l)\frac{\partial e^G}{\partial l} \tag{4}$$

The first term is the same as before; since we have assumed that (i) the perception of gender identity, p, increases as language gender marking l increases, and (ii) gender norms prescribe an effort level e^G for women that is lower than what is optimal in the absence of gender norms (w), it is negative for women and zero or positive for men. The second term reflects the fact that gender norms, and thus the difference in behavioural prescriptions for men and women, are stronger in societies where language gender marking is higher. Therefore, this term, too, is negative for women and positive for men. Consequently, differences in the gender gap in labour supplies across countries will increase in gender marking for two reasons: a stronger gender marking in the local language increases both the gender difference in behavioural prescriptions and the salience of gender as a social category. Empirically, we can no longer say what causes the greater gender gaps in labour supplies, since the effect of language on cognition and the feedback effects between language and social norms are indistinguishable. As will be discussed in more detail in the following section, empiricists have been looking for remedies to mitigate this problem. One potential solution is to study immigrants who live in an institutional context that is different from their home country, and for whom language and institutions therefore do not correlate.

2.5 Incorporating migration

Studying the behaviour of immigrants, however, can introduce new pitfalls to the identification of a causal "cognition" effect of speaking a gendered language on behaviour. To illustrate the issues in our model, we need to add the possibility to migrate between countries with different gender norms: Similar to the AK schooling model, we allow individuals to choose their social category. This may seem unfitting at first, given that the relevant social categories are man and woman, and affiliation with one sex is determined by birth for most people, but we believe this extension will prove useful. While we acknowledge that transgender transitions are also a form of choosing one's gender category, we will not consider such cases here. Rather, we model migration choices as a form of choosing one's gender category: Women (men) may choose whether to live in a society with lower or higher measures of gender equality.

To extend the model in this manner, we need to add another additional component from the AK schooling model. There, identity payoffs are not only determined by students' behaviour, i.e., effort levels, but also by their innate attributes (e.g., looks for the jocks, ability for the nerds). In our application of the model, we include the individual's aptitude in the labour market as an innate attribute. We assume there are two types of individuals who can achieve either high or low returns to effort in the labour market ($w_h > w_l$). We also assume that individuals cannot change the impact

their mother tongue has had on their cognition because the salience of gender for their identity, $p_i(l_i)$, is determined during early childhood and cannot be changed thereafter.[10] Thus, $p_i(l_i)$ is unaffected by the decision to migrate.

We use superscripts to denote high and low gender equality. Thus, l^l (l^h) refers to languages with low (high) gender equality embedded in their grammar, i.e., those that require referring to gender as a category frequently (rarely). Similarly, $e^{(f,l)}$ ($e^{(f,h)}$) refers to the behavioural prescription for women's ideal labour market effort in countries with low (high) gender equality. Since lower gender equality imposes stronger restrictions on women's effort supply, $e^{(f,l)} < e^{(f,h)} \leq w$. For convenience, we assume that women born in countries with a gender-neutral language do not perceive gender identity at all, i.e., $p_i(l_i^h) = 0$. Consequently, these women are indifferent between migrating and staying, regardless of their aptitude. This simplification allows us to focus our attention on women who were born in countries with a gendered language, where, as we discussed earlier, gender norms prescribe higher inequality in men's and women's behaviour. Given these assumptions, and abstracting from moving costs, a woman maximizes her utility by simultaneously choosing her category (woman in a country with high or low gender equality) and effort level:

$$U_i^h = \Big(1 - p_i(l_i^l)\Big)\left[w_i e_i - \frac{1}{2}e_i^2\right] + p_i(l_i^l)\left[-\frac{1}{2}(e_i - e^{f,h}(l^h))^2\right] \tag{5}$$

$$U_i^l = \Big(1 - p_i(l_i^l)\Big)\left[w_i e_i - \frac{1}{2}e_i^2\right] + p_i(l_i^l)\left[-\frac{1}{2}(e_i - e^{f,l}(l^l))^2\right] \tag{6}$$

For certain parameter values of w_l, w_h, $p_i(l_i^h)$, it can be shown that women from countries with lower gender equality gain more from migrating when their returns from supplying effort to the labour market are high.[11] Thus, while women who migrate from countries with high gender equality are equally likely to be of high or low labour market aptitude, women from countries with low gender equality might be positively selected in terms of their aptitude. Thus, on average, the latter group of migrant women might supply even more effort to the labour market by either showing a higher propensity to be employed (extensive margin) or working more hours (intensive margin).

After having illustrated the various challenges in identifying the causal, cognitive effect of language on behaviour, we assess the existing empirical approaches by reviewing

[10] Psycho-linguistic research suggests that gender marking in language influences children's development of gender identity (e.g., Guiora *et al.*, 1982) and that cognitive effects of grammatical gender are most likely formed at an early age (e.g., Flaherty, 2001).

[11] This can be seen by substituting the optimal effort level, $e_i^* = (1-p)w_\theta + pe^G$, with $\theta = h, l$ into the utility function. The resulting function, $U_i(e_i^*) = \frac{1}{2}(1-p)[w_\theta^2 - p(w_\theta - e^G)^2]$, decreases in the absolute deviation between w_θ and e^G. Consequently, when $|w_h - e^{f,l}| > |w_h - e^{f,h}|$, and $|w_l - e^{f,l}| < |w_l - e^{f,h}|$, women with high labour market aptitude will prefer to migrate to a country with higher gender equality, while those with lower aptitude prefer staying.

the economics literature on the subject in the following section.

3 Empirical evidence

For the purpose of the present paper, we will restrict our attention to the evolving literature on the link between gendered language and economic outcomes , and discuss how the evidence it provides allows drawing conclusions about a direct effect of language on behaviour through cognition. The model we presented in Section 2 will serve as a guideline. Appendix-Table 5 provides an overview of the studies discussed here with regard to the data used, the sample restrictions, the way languages are assigned to individuals, and the different measures of gendered language.

3.1 Correlation studies

At the country level, several studies show a correlation between the gender marking of a country's dominant language and different outcomes related to gender (in)equality. Mavisakalyan (2015) and Gay *et al.* (2013) show a negative association between the dominant language's gender marking and women's participation in the labour market. Gay *et al.* (2013) further document a negative relationship with women's representation in politics as well as their access to credit and land. Santacreu-Vasut *et al.* (2013) argue that gender marking in the dominant language should be associated with stark gender imbalances among political representatives, and show that these countries are indeed more likely to introduce gender quotas for the lower house of parliament. In a similar vein, gender marking of the dominant language correlates negatively with female participation in corporate management and with female managers' propensity to lead large teams (Santacreu-Vasut *et al.*, 2014). Jakiela & Ozier (2018) compute fractions of the population that speak gendered languages for a large number of countries and show that a higher value in this measure is associated with higher gender differences in labour market participation and more support for traditional gender roles. For women's educational attainment, they do not find a statistical relationship. Using data from a meta-analysis comparing studies of the gender wage gap, van der Velde *et al.* (2015) show that the gender marking of a country's dominant language is associated with a larger adjusted gender wage gap. Most recently, Hechavarría *et al.* (2017) find that the gender gap in entrepreneurial activity is higher in countries with gendered dominant languages.

3.2 Epidemiological approach

The model discussed in Section 2 illustrates why correlations between gender marking in a country's dominant language and measures of gender inequality at the country level do not permit conclusions regarding the causal impact of language on individual behaviour through cognition. Rather, we view them as evidence that gender norms, language, and institutions co-evolve. The authors of these nonetheless intriguing findings mostly acknowledge this limitation explicitly. In search of a superior identification strategy, suited to providing causal estimates of the cognitive effect of language on individual

behaviour, researchers have turned to the epidemiological approach most prominently applied by Fernández & Fogli (2009). Originally, this approach sought to identify the effect of culture on women's behaviour (fertility and hours worked), while muting the effect of institutions. To this end, it focuses on immigrant women who face the same institutions in one host country but come from different cultural backgrounds. Therefore, the key argument is that one can rule out institutional constraints (e.g., the educational system or availability of child care) and the overall economic environment as influential factors. Any differences in immigrant women's behaviour that are systematically related to outcomes of women in their country of ancestry are instead likely caused by culture (Fernández, 2011).

Using a similar line of reasoning, researchers interested in separating the effect of language on behaviour from the effect of institutions and culture, e.g., Santacreu-Vasut *et al.* (2013), propose using the epidemiological approach to identify the cognition mechanism. To date, most micro-level studies on the relationship between language and economic outcomes follow this recommendation and apply some version of the epidemiological approach (Mavisakalyan & Weber, 2017, provide an overview). However, some differences in the methodology used deserve more attention, such as the type of immigrants sampled (first versus second generation) and assignment of grammatical gender marking to the individual (dominant language of country of origin versus language spoken at home). They turn out to play a crucial role in influencing the issues we raise in this section, which we believe cast doubt on the usefulness of the epidemiological approach for the specific problem at hand: separating a causal cognitive effect of language from the effects of institutions *and* culture.

Before we develop our argument, we note two interesting exceptions among the micro-level studies in this literature. Mavisakalyan (2015) does not study immigrants, but exploits within-country variation in languages spoken, an approach similar to Chen (2013). Using the World Values Survey, she categorizes the language respondents report to speak most frequently at home as highly, mildly, or not gendered. Her main finding is that women speaking a highly gendered language are less likely to participate in the labour market. Moreover, conditional on participation, these women are more likely to be employed part-time. Jakiela & Ozier (2018) apply a very similar strategy to a subsample of the Afrobarometer survey where they use variation in native languages in Kenya, Niger, Nigeria and Uganda. They find that speaking a gendered native language goes along with less labour market participation and lower educational attainment for women relative to men speaking the same language. Investigating within-country variation allows these authors to convincingly separate the effect of language from that of institutions. Notably though, Mavisakalyan (2015) does not interpret her estimates as causal effects of language on behaviour, as even within the same country, she cannot rule out that gender marking in language is correlated with other unobserved characteristics, such as culture or social norms, which could be the true driver of these differences.

Gay et al., in their 2013 working paper, provide results similar to Mavisakalyan

(2015), based on the same data and using the same strategy, but with different indicators for gender marking of the language spoken at home. They show that women speaking highly gendered languages are less likely to be employed and also less likely to work in the agricultural sector. Interestingly, when applying the epidemiological approach to a sample of female immigrants in the U.S., however, they find that, conversely, female immigrants spend more hours in formal employment when reporting that they speak a gendered language at home.[12] The authors argue that the reversal of the effect is due to self-selection into migration, suggesting that the observed women consciously escaped the gender norms of their country of origin. This is consistent with the mechanism we describe in our model: women from countries of origin with low gender equality (which coincide with grammatical gender marking in the language, as the above cited cross-country studies show) migrate at a higher rate when they have a higher aptitude for labour market participation, while women from countries with more gender equality are equally likely to migrate, regardless of their aptitude.

In a more recent paper, the same set of authors, Gay *et al.* (2017), extract a large sample of female immigrants in the U.S. from the American Community Survey (ACS), to which they apply further restriction criteria: the women must (a) be married with a spouse present in the household; and (b) report speaking a language other than English. Among these women, speaking a gendered language is associated with a significantly *lower* propensity to participate in the labour force. Because the results are robust when controlling for linguistic families, even with the inclusion of origin country dummies, the authors interpret this as the cognitive effect of language on behaviour through variation in spoken languages among immigrants from the same country of origin. It is not clear, however, that the epidemiological approach is suited to achieve this. It was designed to separate the effect of culture from institutions (**?**), and was considered superior to comparing individual behaviour across countries while merely controlling for variables meant to capture differences in institutions. The approach loses this advantage when used to study the causal effect of language on behaviour. If language not only affects cognition but also serves as an indicator for deeply rooted social gender norms, then comparing migrants from different cultural backgrounds, albeit within the same institutional setting and incorporating proxies for culture, suffers from the same methodological problem as comparing respondents across countries while including proxies for institutions. Consequently, the results presented by Gay *et al.* (2017) may not permit a "more causal" interpretation than those of Mavisakalyan (2015). Similar concerns apply to the study of Hicks *et al.* (2015), who find that immigrant women in the U.S. from countries with gender marking in the dominant language allocate significantly more time to household chores, while males from these countries report less time allocated to such tasks.

[12] Going back to their World Values Survey sample and restricting it to individuals living in countries where the dominant language is gender-neutral, Gay *et al.* (2013) arrive at a similar conclusion: Women who report speaking a gendered language at home are more likely to be active in the labour market.

3.3 Dominant versus spoken language

Only at first glance does it seem that using the language individuals actually speak at home, rather than the dominant language of the country of origin, may alleviate the concern of assigning culturally ingrained gender norms. Upon deeper reflection, this approach might cause even greater trouble. For example, considering only immigrants who do not speak the host country's dominant language at home (Gay *et al.*, 2017) raises concerns regarding selection: Speaking the host country's language is an indicator for cultural assimilation, which may in turn be related to gender inequality in the country of origin. If women from countries with low gender equality tend to have higher labour market aptitude on average, they might be more eager to integrate into the host country culture. The language individuals speak at home is a matter of choice, and very likely related to other choices, e.g., who to marry. The language spoken at the time of the survey may not be the same language spoken during an individual's critical age period, in which effects on cognition are said to manifest (e.g., Flaherty, 2001; Sera *et al.*, 1994, 2002). Imposing the restriction that a spouse must be present in the household further compounds the problem, as the spouse might act as an enforcer of cultural gender norms picked up by the language. The last column in Table 5 shows how the authors of each paper dealt with matching languages to individuals.

3.4 First versus second generation

Finally, Galor *et al.* (2017) show that female second-generation immigrants in the U.S. are less likely to have attended college when they speak a language with a sex-based grammatical gender system. In the epidemiological literature, studying second-generation immigrants is usually considered superior to studying first-generation immigrants because economic migrants are, on average, more able, aspiring, or in some other way more likely to succeed in the labour market than their otherwise similar counterparts who decided to remain in their home country (Chiswick, 1999). These concerns are typically discounted as minor in samples of second-generation immigrants. However, since empirical studies have shown that cultural beliefs and values are at least partly transmitted from one generation to the next (e.g., Fernández *et al.*, 2004; Farré & Vella, 2013), these individuals are not necessarily better suited to study the causal effect of language on behaviour. If gender marking of language is an indicator for sexist cultural norms, then by assigning the dominant language of the country of origin to second-generation immigrants, we can essentially measure the "intention-to-treat" effect with ancestral culture. Assigning the language most spoken, on the other hand, measures the effect of self-selection into a certain cultural environment within the host country. Additionally, we face the problem of potentially very differently selected first-generation mothers to second-generation daughters. Mothers originating from countries with low gender equality might be more strongly selected in terms of their labour market orientation, or their rejection of the gender norms in their country of origin (Abramitzky *et al.*, 2014). As a result, it seems unlikely that using the epidemiological approach to study

the behaviour of second-generation immigrants can bring us any closer to identifying the causal impact of language.

3.5 Language indicators

Another issue that complicates the comparison of these empirical findings regards the variety of indicators used to measure a language's grammatical gender marking. All studies discussed here rely on four grammatical features relating to gender from the World Atlas of Language Structures (WALS). Most authors (except Mavisakalyan, 2015, see below)[13] dichotomize these features by defining languages as "gendered" vs. "neutral" according to the following rules and justifications:

- Sex-based (SB): The grammatical gender distinction is based on biological sex, as opposed to other distinctions (e.g., animacy) (Corbett, 2013b).
- Number of genders (NG): The language features exactly two genders, as opposed to three or no genders (Corbett, 2013a); the presence of additional gender categories requires less frequent reference to maleness and femaleness (e.g., Hicks *et al.*, 2015).
- Gender assignment (GA): Gender is assigned to nouns according to both semantic and formal rules, as opposed to sematic rules only (Corbett, 2013c); this makes gender more pervasive and therefore more visible (Hicks *et al.*, 2015; Mavisakalyan, 2015).
- Gendered pronouns (GP): The language has gender-specific pronouns in the third person, as well as the first or second person. If a language's gendered pronouns are only in the third person, the literature treats it as gender-neutral (Siewierska, 2013).

Some researchers take these four binary indicators and combine them into a "gender intensity measure" by summing up three or four indicators and, in some instances, conditioning the sum on the language being sex-based (SB = 1). The notable exception is Mavisakalyan (2015), who also relies on WALS data, but uses only the information on gendered pronouns to construct her own indicator. She classifies languages as 'highly gendered' when they feature gender distinctions in third person and first- or second-person singular pronouns, 'mildly gendered' when pronouns are in the third-person singular only, and 'gender-neutral' when neither applies.

In summary, our review of the empirical literature raises several issues that limit the comparability of results across studies. It also raises doubts on the suitability of the epidemiological approach to study the causal effect of gendered language on economic behaviour.

[13] See Column 4 in Table 5 for an overview of language indicators that are utilized by each study.

4 Data and empirical design

In our empirical analysis, we illustrate these unresolved issues with separating the cognitive effect of language from that of cultural factors by considering the impact of the three critical aspects just discussed when assessing the epidemiological approach: (i) the generation of migrants to be studied; (ii) the choice of the language indicator; and (iii) the assignment of languages to individuals.

We exploit the variation in the behaviour of immigrants in Europe who, within one European host country, face the same institutional framework but descend from different cultural backgrounds from various countries of origin. While we acknowledge that the approach is well suited for removing *institutional* constraints as confounding factors, our intention is to highlight the issues raised in the preceding sections related to the intertwining of culture and cognition.. We analyse weekly working hours, thus covering both the extensive and intensive margin of labour market participation, because this outcome has been studied prominently with the epidemiological approach to estimate the causal impact of culture (Fernández & Fogli, 2009).

4.1 Data sources and sample selection

For our empirical investigation, we pool seven cross-sections of the European Social Survey (ESS).[14] The ESS has been conducted bi-annually over a period of 13 years (2002 to 2014) and contains information about migrants in 30 (mostly) European countries, though some waves include respondents in Russia, Turkey, and Israel (ESS, 2018). Our measure of working hours is the self-reported number of hours respondents 'normally work a week (in [their] main job), including any paid or unpaid overtime.' Following the literature, we incorporate the unemployed with zero working hours. We exclude outliers who report working more than 67 hours weekly, but our results are robust to including these observations. Furthermore, only respondents of working age (25 to 55 years) are considered. Appendix-Table 7 presents summary statistics for all relevant variables.

In an attempt to address the empirical issues discussed earlier, of self-selection into migration and intergenerational transmission of culture we split the data set to separately study first-generation (born in a country different from the one where the interview took place) and second-generation (at least one parent born in a foreign country) immigrants.

4.2 Linguistic data

We follow the literature by using the World Atlas of Language Structures (WALS) Online Dryer & Haspelmath (2013) to categorize languages by grammatical gender marking. In order to attain results that are as comprehensive as possible, we present multiple specifications of our regressions, one for each of the four indicators obtained from the WALS as previously described: sex-based (SB), number of genders (NG), gender assignment (GA), and gendered pronouns (GP). A fifth specification features the indicator for the

[14] In the first wave, respondents' spoken languages were not recorded.

number of gendered pronouns used by Mavisakalyan (2015) (GPM). We then present a sixth indicator (GL), which we consider best suited to detect the linguistic incidence of gender marking in a language based on the following considerations: Three of the above four grammatical features—namely, GP, NG, and GA—are only indicative of gender marking in concurrence with grammatical gender categories being based on biological sex (SB), the fourth feature. They do not indicate gender marking if based on some other distinction, such as animate vs. inanimate nouns. Thus, for our GL indicator, we assign gender marking if the language: 1) has a grammatical gender system that is based on biological sex; and 2) is coded as gendered in at least one of the remaining three indicators.

4.3 Language assignment

Another important question is how to assign languages to individuals in the data set. The two strategies employed in the literature use either the dominant language of the country of origin or the self-reported language respondents speak (see last column of Appendix-Table 5 for an overview). Assigning the dominant language is clearly problematic in the case of multilingual countries with no unambiguously dominant language. Consequently, many authors choose the language respondents report to speak most often in daily life. Assigning the self-reported language, however, might contaminate the analysis with endogeneity. First, language use is a choice; it might depend on whether or not a person lives alone, cohabitates with a partner, or resides with family members or compatriots. Thus, we cannot determine whether the language spoken most often is the dominant language in her own country of origin or that of her spouse or another individual living in the residence. The household constellation, in turn, may be affected by both her culturally ingrained gender norms as well as her labour market behaviour. Moreover, speaking the dominant language of the host country or another language is clearly an indicator for the degree of cultural assimilation in the host country, which in turn is an important determinant of labour market outcomes (Chiswick & Miller, 2015). To further explore this issue of possible endogeneity introduced by the spoken language, we present all of our specifications assigning gender marking at both the level of the dominant language of the country of origin and the language most often spoken at home. We compile information on countries' dominant languages from The World Factbook (Central Intelligence Agency (CIA), 2017), the Atlas of the World's Languages (Asher & Moseley, 1994), and the Ethnologue (Gordon, 2005). Appendix-Table 6 provides an overview on dominant languages in the respondents' countries of origin.

4.4 Final dataset and estimated model

Merging all this linguistic information to the ESS data results in a dataset of 7,399 first- and 5,947 second-generation immigrants.[15] Appendix-Table 7 presents summary

[15]Because the number of second-generation immigrants in the ESS is rather limited, we also include immigrants of the so-called "1.5th generation," i.e., those that arrived in the host country together with their parents before the age of 14. For those individuals, it seems safe to assume that the decision to

statistics for the two different datasets.

With this data, we estimate the following model using ordinary least squares (OLS) regressions:

$$Hours_{icklt} = \beta_0 + \beta_1 GL_l + \beta_2 \boldsymbol{Y}_k + \beta_3 \boldsymbol{X}_i + \beta_4 \boldsymbol{\delta}_c + \beta_5 \boldsymbol{\theta}_t + \epsilon_{icklt} \tag{7}$$

$Hours_{icklt}$ represents the weekly working hours of individual i living in host country c, descending from ancestry k, with dominant language/speaking language l, who is observed at time period t. GL is a binary variable, taking on the value of 1 when the respondent's language features gender marking, and 0 otherwise. $\boldsymbol{Y}$ is a vector of characteristics of the respondent's country of origin used in the literature to capture variation in cultural attitudes and norms: The rate of female labour force participation (FLFP), taken from the International Labour Organization (ILO, 2018), total fertility rates (TFR) from the World Bank (2018), and GDP per capita from the United Nations Statistics Division (UN, 2018). $\boldsymbol{X}$ represents a vector of controls at the individual level: age, age squared, educational level of the respondent's mother (one dummy variable for lower than secondary education and one for tertiary education, leaving secondary education as the reference group), and religion (dummy variables for Christian and Muslim faith, as well as Eastern religions, leaving respondents not reporting any affiliation with a religious group as the reference group). $\boldsymbol{\delta}$ is a set of dummy variables for the host countries, which we include to account for heterogeneous institutions and economic conditions in locations across European destinations. Finally, $\boldsymbol{\theta}$ represents a set of time dummies for the survey waves. As the working hours of women and men are estimated within the same model[16], a female dummy and all interactions of the female dummy with the country of origin characteristics and host country dummies are also included in the estimation (but not all are displayed, results available upon request).

5 Results

The estimated coefficients for the model are displayed in Tables 1 and 2 for first-generation, and Tables 3 and 4 for second-generation immigrants. One general finding is that the coefficient estimate for the female dummy is always negative and large in size, indicating that, among migrants with a language background not marked by gender, women work fewer hours than men, though this finding is only statistically significant among second-generation immigrants, as the standard errors are too large in the first-generation sample. A second general finding is that, compared to the baseline category of respondents whose mother acquired secondary education, lower maternal education reduces respondents' working hours significantly. Mothers' education is particularly in-

migrate and the selection of the host country was their parents' decision and thus exogenous to the respondents, as in the case of second-generation immigrants.

[16] We estimate joint regressions to enhance readability of the already extensive tables. We also estimated the models separately for females and males with largely unchanged results (available upon request).

fluential in terms of the labour force activity of first-generation immigrants, for whom the coefficient estimate is twice as large as for second-generation immigrants. Similarly, the finding that respondents of Muslim faith work significantly fewer hours than all other religious (or atheist) groups is more pronounced among the first generation than the second.

FLFP in the ancestry country is negatively related to men's current working hours in both generations. A significantly positive effect on women's labour supply is observed for first-generation immigrants only and may be interpreted as a selection effect. The two samples are differentially affected by the ancestry country's TFR. Whereas higher TFR decreases men's and women's working hours among first-generation immigrants, a significantly and substantially negative impact persists only for women in the second generation. The relationship between GDP and working hours is U-shaped for second-generation immigrants. By and large, the estimation results confirm the expectation of a more diverse sample of first-generation immigrants, whose working hours overall seem more responsive to the circumstances in their origin countries than the subsequent generation's. Nonetheless, the labour market behaviour of the second generation still clearly relates to their parents' educational, religious, and economic origins.

Finally, and most importantly for our research question, the estimation analysis reveals that none of the language indicators for either spoken or dominant language is systematically related to the working hours of men. The interaction between language gender marking and being female, however, shows an economically large and statistically significant coefficient for some of the indicators.

To facilitate interpretation, we plot the marginal effects for females (f) and males (m), respectively, in Figure 1. The top row refers to first, the bottom row to second generation immigrants. Within each row, Panels 1 and 2 show the effects on working hours obtained by assigning the spoken or dominant language. The figures show that language gender marking in the spoken or dominant language is hardly related to the working hours of women either. If any, as for the GA indicator in the bottom left panel of Figure 1, the effect on hours is positive, meaning that a woman whose dominant language in her parents' country of origin assigns nouns to genders on both semantic and formal rules is observed to work, on average, 1.5 more hours per week than a comparable woman from an ancestry country with a language that assigns gender only on semantic grounds or not at all. The GL indicator has a coefficient estimate of similar size, meaning that a woman's working hours increase by about the same number when originating from a gendered dominant language country, though the effect is only borderline significant. A general observation is that none of the estimated indicator coefficients supports a negative impact of speaking a gendered language on women's labour market activity. These findings are in line with the implications derived from our model in Section 2: if country-of-origin language reflects the local cultural values and gender norms, women who stand to gain from rejecting these norms will select into migration at a higher rate than others. Focusing on second-generation immigrants does not seem to solve

this problem, as the relevant cultural attitudes may still be transmitted from parents to their children. Another general observation is that these coefficients, though not statistically significantly different from zero on average, have slightly larger variance among first-generation immigrants than second-generation immigrants, whose working hours are likely more homogenous after having experienced a longer integration period.

Overall, our results confirm the pitfalls of trying to identify causal effects of language on behaviour with the epidemiological approach laid out in the conceptual framework: Because culture and language are intertwined, studying immigrants from heterogeneous language backgrounds does not help separate the causal impact of language. Additionally, in this context, the epidemiological approach seems to suffer from (parental) selection issues in both generations of immigrants.

6 Conclusion

Our empirical finding of negligible effects of the language indicators are fully in line with the arguments we derived from our theoretical framework and from a systematic review of the empirical literature. Based on (psycho-)linguistic and identity economics considerations, the basic version of our model shows that the extent to which people conform with their socially prescribed roles depends on the gender marking of their language, i.e., the emphasis it places on the category of gender. The extended version of the model incorporates the interrelatedness of language and culture, accounting for feedback effects of increased gender marking on the gender differences in behavioural prescriptions. As a result, culture, norms, and language effects cannot be easily discriminated empirically.

While the widely used epidemiological approach addresses the co-evolvement of culture and institutions, it cannot resolve the interlocking of culture and language. Moreover, in the present context, it does not seem suited to overcome the issue of self-selection and parental selection, which we carved out in a further extension of our model, in which women may choose a country of residence with high or low gender equality and their (socially desired) work effort level simultaneously. One important insight from this model was that gender marking may affect individual behaviour through two channels: (a) grammatical structure by way of cognition; and/or (b) institutions and cultural norms. While not mutually exclusive, these two channels present a challenge for empiricists trying to measure the effect of language on behaviour and establish a causal link.

In our assessment of the empirical literature, we illustrated that the two channels are indeed not easily discriminated. As a consequence, the gender marking of an individual's language does not seem to have a systematic and robust impact on labour market behaviour. None of the language indicators in our application of the epidemiological approach, neither for spoken nor for dominant language, is systematically related to the working hours of immigrant men; this is true for both first- and second-generation immigrant men. Second-generation immigrant women's working hours are related to only one aspect of gender marking of the dominant language: whether nouns are assigned to genders on both semantic and formal rules. Consequently, our results do not confirm

Figure 1: Estimated marginal effects of gendered language on working hours

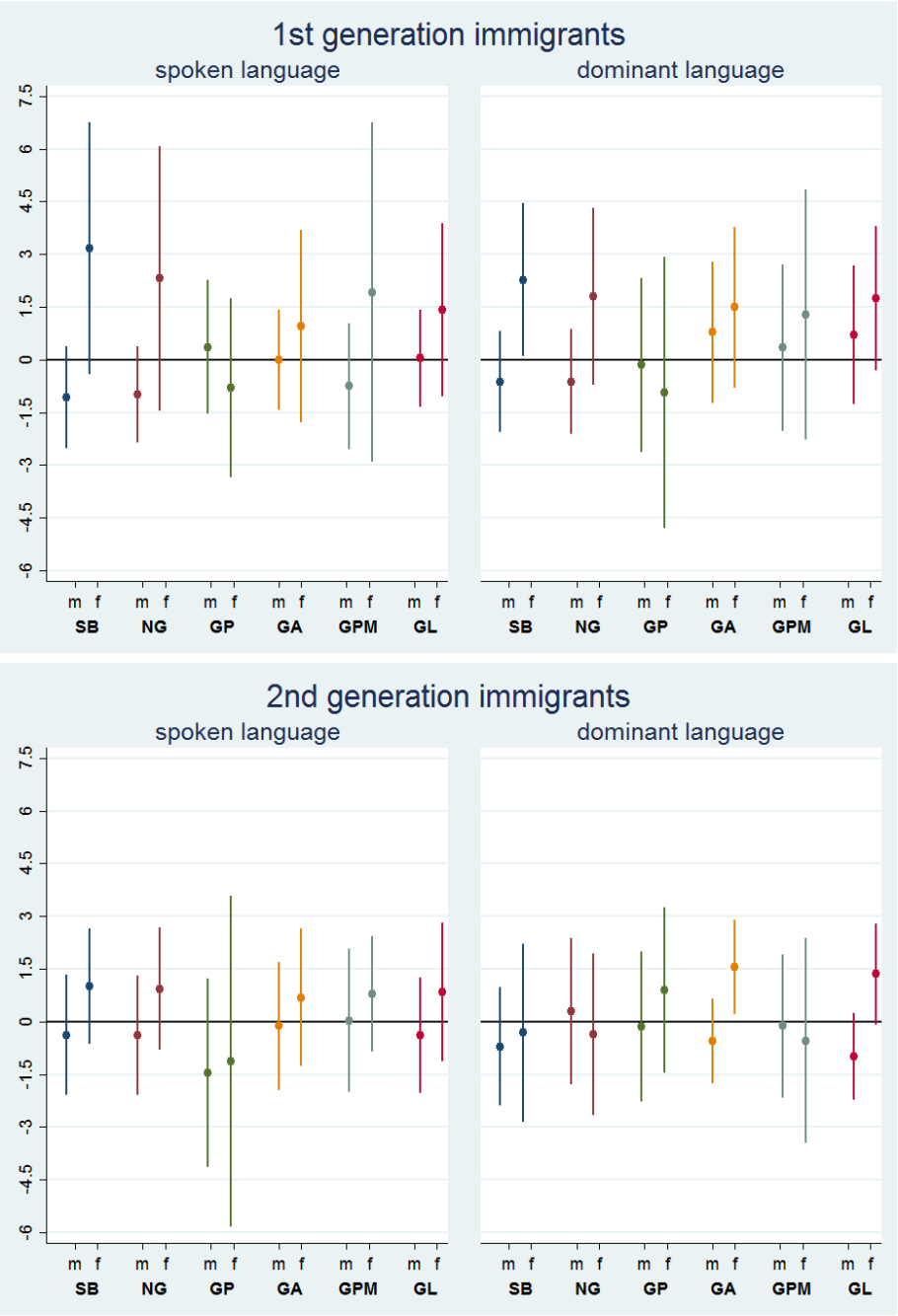

Table 1: Regression results for first generation immigrants, spoken language

	SB	NG	GP	GA	GPM	GL
			Individual characteristics			
Female	-13.186	-13.101	-11.181	-13.079	-12.079	-13.332
	(11.967)	(12.276)	(12.027)	(12.374)	(12.324)	(12.107)
Age	0.358	0.357	0.362*	0.365*	0.361*	0.368*
	(0.216)	(0.216)	(0.216)	(0.215)	(0.216)	(0.215)
Age squared	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Mother no education[a]	-1.857***	-1.871***	-1.872***	-1.873***	-1.875***	-1.870***
	(0.404)	(0.407)	(0.409)	(0.407)	(0.411)	(0.405)
Mother tertiary education	1.691***	1.697***	1.703***	1.709***	1.706***	1.708***
	(0.355)	(0.354)	(0.354)	(0.354)	(0.353)	(0.356)
Christian	-0.253	-0.237	-0.229	-0.255	-0.227	-0.271
	(0.314)	(0.313)	(0.307)	(0.317)	(0.310)	(0.319)
Eastern	-0.188	-0.239	-0.348	-0.245	-0.317	-0.200
	(0.846)	(0.858)	(0.853)	(0.857)	(0.862)	(0.851)
Islamic	-4.068***	-4.157***	-4.218***	-4.094***	-4.200***	-4.021***
	(0.746)	(0.713)	(0.613)	(0.687)	(0.640)	(0.717)
			Characteristics of the ancestry country			
FLFP	-7.843***	-7.922***	-7.556**	-7.674**	-7.986**	-7.569**
	(2.949)	(2.950)	(3.064)	(2.972)	(3.088)	(2.928)
FLFP x female	10.606**	10.568**	9.797**	10.939**	10.695**	11.311**
	(4.178)	(4.214)	(4.418)	(4.484)	(4.522)	(4.453)
TFR	-0.733**	-0.723**	-0.697*	-0.702*	-0.698*	-0.703**
	(0.355)	(0.357)	(0.364)	(0.356)	(0.362)	(0.347)
TFR x female	0.068	0.055	0.023	0.161	0.006	0.222
	(0.599)	(0.603)	(0.625)	(0.655)	(0.627)	(0.642)
GDP	-0.200	-0.206	-0.210	-0.215	-0.221	-0.212
	(0.198)	(0.198)	(0.198)	(0.181)	(0.199)	(0.184)
GDP x female	0.065	0.088	0.103	0.184	0.123	0.186
	(0.193)	(0.196)	(0.196)	(0.203)	(0.194)	(0.194)
GDP squared	0.006	0.006	0.006	0.007	0.007	0.007
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
GDP squared x female	0.002	0.001	-0.000	-0.003	-0.001	-0.003
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
			Language indicator (LI[b])			
LI	-1.081	-0.997	0.361	0.001	-0.291	0.040
	(0.725)	(0.688)	(0.954)	(0.717)	(0.663)	(0.692)
LI x female	3.173*	2.313	-0.793	0.955	0.511	1.429
	(1.799)	(1.888)	(1.273)	(1.373)	(1.502)	(1.234)
Host country dummies[c]	YES	YES	YES	YES	YES	YES
Year dummies	YES	YES	YES	YES	YES	YES
Constant	42.963***	42.903***	41.641***	41.674***	42.169***	41.517***
	(5.396)	(5.396)	(5.282)	(5.297)	(5.313)	(5.290)
Observations	7,399	7,399	7,399	7,399	7,399	7,399
Adjusted R-squared	0.150	0.150	0.150	0.150	0.150	0.150

Note: Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

[a] Includes mothers with primary education.
[b] From left to right, $LI = SB, NG, GP, GA, GPM, GL$.
[c] All interactions with *Female* are also included.

Table 2: Regression results for first generation immigrants, dominant language

	SB	NG	GP	GA	GPM	GL
			Individual characteristics			
Female	-13.903	-13.850	-10.656	-15.659	-12.154	-15.818
	(12.107)	(12.273)	(11.867)	(12.111)	(12.015)	(11.927)
Age	0.362*	0.361*	0.363*	0.383*	0.361*	0.389*
	(0.217)	(0.216)	(0.216)	(0.214)	(0.215)	(0.215)
Age squared	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Mother no education[a]	-1.829***	-1.853***	-1.888***	-1.819***	-1.863***	-1.790***
	(0.414)	(0.415)	(0.417)	(0.398)	(0.417)	(0.401)
Mother tertiary education	1.679***	1.685***	1.717***	1.674***	1.698***	1.667***
	(0.353)	(0.354)	(0.354)	(0.358)	(0.354)	(0.358)
Christian	-0.254	-0.247	-0.230	-0.265	-0.232	-0.278
	(0.311)	(0.310)	(0.310)	(0.311)	(0.307)	(0.313)
Eastern	-0.040	-0.118	-0.431	0.085	-0.219	0.040
	(0.830)	(0.844)	(0.884)	(0.839)	(0.889)	(0.814)
Islamic	-4.073***	-4.139***	-4.219***	-3.793***	-4.172***	-3.728***
	(0.732)	(0.699)	(0.577)	(0.765)	(0.650)	(0.789)
			Characteristics of the ancestry country			
FLFP	-8.145**	-8.237**	-8.023**	-6.047*	-7.739**	-5.896*
	(3.231)	(3.242)	(3.831)	(3.378)	(3.544)	(3.497)
FLFP x female	11.624**	11.494**	8.864*	12.535**	10.796**	13.518***
	(4.528)	(4.551)	(5.295)	(4.866)	(5.047)	(4.973)
TFR	-0.792**	-0.787*	-0.679*	-0.486	-0.695*	-0.511
	(0.393)	(0.395)	(0.373)	(0.427)	(0.366)	(0.424)
TFR x female	0.261	0.217	0.060	0.481	0.013	0.578
	(0.680)	(0.672)	(0.636)	(0.702)	(0.635)	(0.705)
GDP	-0.207	-0.214	-0.222	-0.096	-0.214	-0.137
	(0.197)	(0.198)	(0.206)	(0.179)	(0.206)	(0.164)
GDP x female	0.057	0.087	0.088	0.285	0.122	0.255
	(0.210)	(0.208)	(0.194)	(0.220)	(0.197)	(0.195)
GDP squared	0.006	0.007	0.007	0.003	0.007	0.004
	(0.007)	(0.007)	(0.008)	(0.006)	(0.008)	(0.006)
GDP squared x female	0.002	0.001	0.000	-0.006	-0.001	-0.004
	(0.008)	(0.008)	(0.007)	(0.008)	(0.007)	(0.007)
			Language indicator (LI[b])			
LI	-0.624	-0.622	-0.154	0.777	0.008	0.717
	(0.725)	(0.750)	(1.244)	(1.005)	(0.691)	(0.988)
LI x female	2.279**	1.805	-0.940	1.490	0.278	1.755*
	(1.087)	(1.258)	(1.938)	(1.145)	(1.175)	(1.029)
Host country dummies[c]	YES	YES	YES	YES	YES	YES
Year dummies	YES	YES	YES	YES	YES	YES
Constant	42.679***	42.737***	41.851***	38.905***	41.732***	38.937***
	(5.660)	(5.649)	(5.401)	(5.816)	(5.552)	(5.812)
Observations	7,399	7,399	7,399	7,399	7,399	7,399
Adjusted R-squared	0.150	0.150	0.150	0.151	0.150	0.151

Note: Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

[a] Includes mothers with primary education.
[b] From left to right, $LI = SB, NG, GP, GA, GPM, GL$.
[c] All interactions with *Female* are also included.

Table 3: Regression results for second generation immigrants, spoken language

	SB	NG	GP	GA	GPM	GL
			Individual characteristics			
Female	-13.671**	-13.670**	-13.464**	-13.732**	-13.700**	-13.755**
	(5.463)	(5.463)	(5.446)	(5.464)	(5.451)	(5.455)
Age	0.290*	0.291*	0.301*	0.289*	0.294*	0.291*
	(0.170)	(0.170)	(0.170)	(0.169)	(0.172)	(0.170)
Age squared	-0.003	-0.003	-0.003*	-0.003	-0.003	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Mother no education[a]	-0.870**	-0.871**	-0.877**	-0.871**	-0.882**	-0.874**
	(0.415)	(0.415)	(0.406)	(0.416)	(0.418)	(0.415)
Mother tertiary education	0.183	0.183	0.143	0.182	0.176	0.183
	(0.466)	(0.466)	(0.463)	(0.467)	(0.465)	(0.466)
Christian	0.119	0.123	0.131	0.117	0.151	0.128
	(0.327)	(0.326)	(0.330)	(0.334)	(0.324)	(0.335)
Eastern	0.514	0.509	0.447	0.503	0.459	0.493
	-1.632	-1.631	-1.645	-1.635	-1.612	-1.637
Islamic	-2.921***	-2.926***	-2.835***	-2.897***	-2.965***	-2.934***
	(1.096)	(1.096)	(1.065)	(1.089)	(1.068)	(1.086)
			Characteristics of the ancestry country			
FLFP	-4.928*	-4.931*	-5.172**	-4.908*	-5.070**	-4.975*
	(2.509)	(2.510)	(2.526)	(2.500)	(2.494)	(2.506)
FLFP x female	4.043	4.039	3.733	4.107	4.072	4.140
	(3.600)	(3.597)	(3.492)	(3.579)	(3.527)	(3.575)
TFR	-0.007	-0.007	-0.004	-0.003	-0.009	-0.011
	(0.303)	(0.303)	(0.302)	(0.306)	(0.304)	(0.306)
TFR x female	-1.330**	-1.331**	-1.363**	-1.325*	-1.341**	-1.320*
	(0.657)	(0.658)	(0.653)	(0.666)	(0.655)	(0.665)
GDP	-0.173	-0.173	-0.190*	-0.174	-0.183	-0.175
	(0.110)	(0.110)	(0.113)	(0.112)	(0.113)	(0.111)
GDP x female	-0.466*	-0.465*	-0.468*	-0.456*	-0.457*	-0.459*
	(0.246)	(0.246)	(0.243)	(0.248)	(0.246)	(0.247)
GDP squared	0.006*	0.006*	0.007*	0.006*	0.006*	0.006*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
GDP squared x female	0.013	0.013	0.013	0.012	0.012	0.012
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
			Language indicator (LI[b])			
LI	-0.376	-0.387	-1.456	-0.117	-0.501	-0.394
	(0.859)	(0.852)	(1.339)	(0.912)	(0.745)	(0.821)
LI x female	1.015	0.933	-1.118	0.685	0.316	0.842
	(0.823)	(0.868)	(2.358)	(0.977)	(1.109)	(0.990)
Host country dummies[c]	YES	YES	YES	YES	YES	YES
Year dummies	YES	YES	YES	YES	YES	YES
Constant	42.175***	42.180***	41.852***	41.944***	42.336***	42.212***
	(3.408)	(3.407)	(3.396)	(3.446)	(3.383)	(3.402)
Observations	5,947	5,947	5,947	5,947	5,947	5,947
Adjusted R-squared	0.124	0.124	0.124	0.124	0.124	0.124

Note: Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

[a] Includes mothers with primary education.
[b] From left to right, $LI = SB, NG, GP, GA, GPM, GL$.
[c] All interactions with *Female* are also included.

Table 4: Regression results for second generation immigrants, dominant language

	SB	NG	GP	GA	GPM	GL
			Individual characteristics			
Female	-13.294**	-13.079**	-14.180**	-17.109***	-14.160**	-16.803***
	(5.488)	(5.305)	(5.537)	(5.489)	(5.459)	(5.514)
Age	0.292*	0.291*	0.291*	0.291*	0.290*	0.295*
	(0.167)	(0.168)	(0.168)	(0.168)	(0.168)	(0.168)
Age squared	-0.003*	-0.003*	-0.003*	-0.003*	-0.003*	-0.003*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Mother no education[a]	-0.908**	-0.868**	-0.867**	-0.873**	-0.869**	-0.887**
	(0.406)	(0.407)	(0.413)	(0.414)	(0.410)	(0.412)
Mother tertiary education	0.201	0.176	0.175	0.195	0.176	0.191
	(0.475)	(0.470)	(0.467)	(0.466)	(0.470)	(0.467)
Christian	0.156	0.131	0.128	0.117	0.127	0.139
	(0.326)	(0.330)	(0.332)	(0.334)	(0.331)	(0.329)
Eastern	0.455	0.469	0.523	0.533	0.524	0.533
	(1.686)	(1.650)	(1.677)	(1.632)	(1.669)	(1.650)
Islamic	-3.193**	-2.883**	-2.900***	-2.857**	-2.895**	-3.040**
	(1.259)	(1.173)	(1.060)	(1.148)	(1.121)	(1.199)
			Characteristics of the ancestry country			
FLFP	-5.481**	-4.667*	-5.086*	-5.576**	-5.007*	-6.573**
	(2.692)	(2.604)	(3.030)	(2.770)	(2.730)	(2.917)
FLFP x female	3.827	3.697	5.375	6.162*	4.520	6.293*
	(3.687)	(3.687)	(3.793)	(3.632)	(3.594)	(3.726)
TFR	-0.048	0.033	0.015	-0.120	0.004	-0.227
	(0.308)	(0.287)	(0.319)	(0.309)	(0.305)	(0.310)
TFR x female	-1.372**	-1.390**	-1.434**	-0.984	-1.357**	-1.016
	(0.646)	(0.644)	(0.644)	(0.669)	(0.662)	(0.643)
GDP	-0.148	-0.183	-0.178	-0.210*	-0.174	-0.212*
	(0.127)	(0.127)	(0.110)	(0.111)	(0.108)	(0.111)
GDP x female	-0.452*	-0.453*	-0.426	-0.333	-0.456*	-0.404*
	(0.252)	(0.248)	(0.267)	(0.245)	(0.249)	(0.237)
GDP squared	0.005	0.007	0.006*	0.007**	0.006*	0.007*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
GDP squared x female	0.012	0.012	0.011	0.008	0.012	0.011
	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)
			Language indicator (LI[b])			
LI	-0.713	0.289	-0.133	-0.543	-0.069	-0.979
	(0.843)	(1.044)	(1.071)	(0.601)	(0.540)	(0.617)
LI x female	-0.317	-0.358	0.906	1.545**	0.305	1.352*
	(1.264)	(1.154)	(1.177)	(0.670)	(0.449)	(0.724)
Host country dummies[c]	YES	YES	YES	YES	YES	YES
Year dummies	YES	YES	YES	YES	YES	YES
Constant	42.719***	41.389***	41.840***	42.919***	41.887***	43.957***
	(3.628)	(3.654)	(3.546)	(3.562)	(3.604)	(3.566)
Observations	5,947	5,947	5,947	5,947	5,947	5,947
Adjusted R-squared	0.124	0.124	0.124	0.124	0.124	0.124

Note: Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

[a] Includes mothers with primary education.
[b] From left to right, $LI = SB, NG, GP, GA, GPM, GL$.
[c] All interactions with *Female* are also included.

the leading hypothesis in the literature, which is that behaviour is more consistent with gender stereotypes for speakers of a gendered language. Instead, behaviour seems hardly related to speaking a gendered language and, if anything, women work more hours. We argue that these results support the idea that language reflects cultural traits, and that men and women who select into migration as first-generation immigrants are more likely to reject the culture in their country of origin, and may transmit this attitude to their children (second generation). With this paper, we hope to have contributed to a systematic assessment by shedding both light and doubt on the growing literature aiming to investigate a causal link between gendered language and gender gaps in various economic outcomes.

References

Abramitzky, Ran, Boustan, Leah Platt, & Eriksson, Katherine. 2014. A nation of immigrants: Assimilation and economic outcomes in the age of mass migration. *Journal of Political Economy*, **122**(3), 467–506.

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

Akerlof, George A, & Kranton, Rachel E. 2002. Identity and schooling: Some lessons for the economics of education. *Journal of Economic Literature*, 1167–1201.

Asher, Ron E, & Moseley, Christopher. 1994. *Atlas of the world's languages*. Routledge.

Becker, Gary S. 1985. Human capital, effort, and the sexual division of labor. *Journal of Labor Economics*, **3**(1, Part 2), S33–S58.

Beit-Hallahmi, Benjamin, Catford, John C, Cooley, Ralph E, Dull, Cecelia Yoder, Guiora, Alexander Z, & Raluszny, Maria. 1974. Grammatical gender and gender identity development: Cross cultural and cross lingual implications. *American journal of orthopsychiatry*, **44**(3), 424.

Beller, Sieghard, Brattebø, Karen Fadnes, Lavik, Kristina Osland, Reigstad, Rakel Drønen, & Bender, Andrea. 2015. Culture or language: what drives effects of grammatical gender? *Cognitive linguistics*, **26**(2), 331–359.

Boroditsky, Lera, Schmidt, Lauren A., & Phillips, Webb. 2003. Sex, syntax, and semantics. *Pages 61–79 of:* Genter, Dedre, & Goldin-Meadow, Susan (eds), *Language in mind: Advances in the study of language and thought*. MIT.

Casasanto, Daniel. 2016. Linguistic relativity. *Pages 158–174 of:* Riemer, Nick (ed), *The routledge handbook of semantics*. Routledge.

Casasanto, Daniel, Boroditsky, Lera, Phillips, Webb, Greene, Jesse, Goswami, Shima, Bocanegra-Thiel, Simon, Santiago-Diaz, Ilia, Fotokopoulu, Olga, Pita, Ria, & Gil, David. 2004. How deep are effects of language on thought? time estimation in speakers of english, indonesian, greek, and spanish. *In: Proceedings of the cognitive science society*, vol. 26.

Chen, M. Keith. 2013. The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. *American Economic Review*, **103**(2), 690–731.

Chiswick, Barry R. 1999. Are immigrants favorably self-selected? *American economic review*, **89**(2), 181–185.

Chiswick, Barry R, & Miller, Paul W. 2015. International migration and the economics of language. *Pages 211–269 of:* Chiswick, Barry R, & Miller, Paul W (eds), *Handbook of the economics of international migration*, vol. 1. Elsevier.

Corbett, Greville G. 2013a. Number of genders. *In:* Dryer, Matthew S., & Haspelmath, Martin (eds), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Corbett, Greville G. 2013b. Sex-based and non-sex-based gender systems. *In:* Dryer, Matthew S., & Haspelmath, Martin (eds), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Corbett, Greville G. 2013c. Systems of gender assignment. *In:* Dryer, Matthew S., & Haspelmath, Martin (eds), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Dryer, Matthew S., & Haspelmath, Martin (eds). 2013. *Wals online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Duflo, Esther. 2012. Women empowerment and economic development. *Journal of economic literature*, **50**(4), 1051–1079.

Farré, Lídia, & Vella, Francis. 2013. The intergenerational transmission of gender role attitudes and its implications for female labour force participation. *Economica*, **80**(318), 219–247.

Fernández, Raquel. 2011. Does culture matter? *In:* Benhabib, J., Jackson M., & Bisin, A. (eds), *Handbook of social economics*, vol. Vol. 1A. North-Holland.

Fernández, Raquel, & Fogli, Alessandra. 2009. Culture: An empirical investigation of beliefs, work, and fertility. *American economic journal: Macroeconomics*, **1**(1), 146–177.

Fernández, Raquel, Fogli, Alessandra, & Olivetti, Claudia. 2004. Mothers and sons: Preference formation and female labor force dynamics. *The quarterly journal of economics*, 1249–1299.

Flaherty, Mary. 2001. How a language gender system creeps into perception. *Journal of cross-cultural psychology*, **32**(1), 18–31.

Fortin, Nicole M. 2005. Gender role attitudes and the labour-market outcomes of women across OECD countries. *Oxford review of economic policy*, **21**(3), 416–438.

Galor, Oded, Özak, Ömer, & Sarid, Assaf. 2017. *Geographical origins and economic consequences of language structures*. Ssrn working paper. Available online: https://ssrn.com/abstract=2820889orhttp://dx.doi.org/10.2139/ssrn.2820889 (accessed May 18, 2018).

Gay, Victor, Santacreu-Vasut, Estefania, & Shoham, Amir. 2013. *The grammatical origins of gender roles*. Working paper. Berkeley Economic History Lab. Available online: http://behl.berkeley.edu/working-papers/ (accessed May 18, 2018).

Gay, Victor, Hicks, Daniel L., Santacreu-Vasut, Estefania, & Shoham, Amir. 2017. Decomposing culture: an analysis of gender, language, and labor supply in the household. *Review of Economics of the Household*, Apr.

Givati, Yehonatan, & Troiano, Ugo. 2012. Law, economics, and culture: Theory of mandated benefits and evidence from maternity leave policies. *The journal of law and economics*, **55**(2), 339–364.

Gordon, R. G. (ed). 2005. *Ethnologue: Languages of the world*. SIL International.

Guiora, Alexander Z, Beit-Hallahmi, Benjamin, Fried, Risto, & Yoder, Cecelia. 1982. Language environment and gender identity attainment. *Language learning*, **32**(2), 289–304.

Hechavarría, Diana M, Terjesen, Siri A, Stenholm, Pekka, Brännback, Malin, & Lång, Stefan. 2017. More than words: do gendered linguistic structures widen the gender gap in entrepreneurial activity? *Entrepreneurship theory and practice*.

Hicks, Daniel L, Santacreu-Vasut, Estefania, & Shoham, Amir. 2015. Does mother tongue make for women's work? Linguistics, household labor, and gender identity. *Journal of economic behavior & organization*, **110**, 19–44.

Hunt, Earl, & Agnoli, Franca. 1991. The Whorfian hypothesis: A cognitive psychology perspective. *Psychological review*, **98**(3), 377.

ILO. 2018. *ILOSTAT: Labour force participation rate by sex and age (%)*. http://www.ilo.org/ilostat/faces/oracle/webcenter/portalapp/pagehierarchy/Page27.jspx?subject=EAP&indicator=EAP_DWAP_SEX_AGE_RT&datasetCode=A&collectionCode=YI&_afrLoop=633361233136369&_afrWindowMode=0&_afrWindowId=t4u2a38my_1#!%40%40%3Findicator%3DEAP_DWAP_SEX_AGE_RT%26_afrWindowId%3Dt4u2a38my_1%26subject%3DEAP%26_afrLoop%3D633361233136369%26datasetCode%3DA%26collectionCode%3DYI%26_afrWindowMode%3D0%26_adf.ctrl-state%3Dt4u2a38my_66 (accessed May 18, 2018).

Jakiela, Pamela, & Ozier, Owen. 2018. *Gendered language*. Tech. rept.

Konishi, Toshi. 1993. The semantics of grammatical gender: A cross-cultural study. *Journal of psycholinguistic research*, **22**(5), 519–534.

MacKay, Donald G. 1999. *Gender in english, german, and other languages: Problems with the old theory, opportunities for the new*. Wiesbaden: VS Verlag für Sozialwissenschaften. Pages 73–87.

Mavisakalyan, Astghik. 2015. Gender in language and gender in employment. *Oxford development studies*, **43**(4), 403–424.

Mavisakalyan, Astghik, & Weber, Clas. 2017. Linguistic structures and economic outcomes. *Journal of economic surveys*.

OECD. 2019. *OECD.Stat: Employment ratio by sex and age group*. https://stats.oecd.org/index.aspx?queryid=54742# (accessed April 11, 2019).

Roberts, Seán, & Winters, James. 2013. Linguistic diversity and traffic accidents: Lessons from statistical studies of cultural traits. *PloS one*, **8**(8), e70902.

Roberts, Seán G, Winters, James, & Chen, Keith. 2015. Future tense and economic decisions: controlling for cultural evolution. *Plos one*, **10**(7), e0132145.

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

Santacreu-Vasut, Estefania, Shenkar, Oded, & Shoham, Amir. 2014. Linguistic gender marking and its international business ramifications. *Journal of international business studies*, **45**(9), 1170–1178.

Sera, Maria D, Berge, Christian AH, & del Castillo Pintado, Javier. 1994. Grammatical and conceptual forces in the attribution of gender by English and Spanish speakers. *Cognitive development*, **9**(3), 261–292.

Sera, Maria D, Elieff, Chryle, Forbes, James, Burch, Melissa Clark, Rodríguez, Wanda, & Dubois, Diane Poulin. 2002. When language affects cognition and when it does not: An analysis of grammatical gender and classification. *Journal of experimental psychology: General*, **131**(3), 377.

Siewierska, A. 2013. Gender distinctions in independent personal pronouns. ms dryer & m. haspelmath (eds.) the world atlas of language structures online. leipzig: Max planck institute for evolutionary anthropology. *In:* Dryer, Matthew S., & Haspelmath, Martin (eds), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Slobin, Dan I. 1987. Thinking for speaking. *Pages 435–445 of: Annual meeting of the berkeley linguistics society*, vol. 13.

Slobin, Dan I. 1996. From "thought and language" to "thinking for speaking.". *Pages 70–97 of:* Gumperz, J.J., & Levinson, S.C. (eds), *Rethinking linguistic relativity*. Cambridge University Press.

Slobin, Dan I. 2003. Language and thought online: Cognitive consequences of linguistic relativity. *Pages 157–191 of:* Gentner, D., & Goldin-Meadow, S. (eds), *Language in mind: Advances in the study of language and thought*. MIT Press.

Stahlberg, Dagmar, Braun, Friederike, Irmen, Lisa, & Sczesny, Sabine. 2007. *Representation of the sexes in language*. Pages 163–187.

Tabellini, Guido. 2008. The scope of cooperation: Values and incentives. *The quarterly journal of economics*, **123**(3), 905–950.

Tainio, Liisa. 2006. Gender in finnish language use: Equal, inequal and/or queer. *Webfu: Wiener elektronische beiträge des instituts für finno-ugristik*, **10**, 1–19.

UN. 2018. *Un data: Per capita gdp at current prices – us dollars.* http://data.un.org/Data.aspx?d=SNAAMA&f=grID%3A101%3BcurrID%3AUSD%3BpcFlag%3A1 (accessed May 18, 2018).

UNDP. 2015. *Human development report 2015. work for human development.*

van der Velde, Lucas, Tyrowicz, Joanna, & Siwinska, Joanna. 2015. Language and (the estimates of) the gender wage gap. *Economics letters*, **136**, 165–170.

Whorf, Benjamin Lee. 1956. *Language, thought, and reality: Selected writings of benjamin lee whorf*. MASS. Edited by John B. Caroll.

World Bank. 2018. *The World Bank Data Base: Fertility rates, total (births per woman).* https://data.worldbank.org/indicator/SP.DYN.TFRT.IN? (accessed May 18, 2018).

7 Appendix

Table 6: Sample countries, languages and language indicators

Country of origin	Dominant language	N	SB	NG	GA	GP	GPM	GL
Albania	Albanian	231	1	1	1	0	1	1
Algeria	Arabic	247	1	1	1	1	2	1
Argentina	Spanish	116	1	1	1	1	2	1
Australia	English	79	1	1	0	0	1	0
Austria	German	147	1	1	1	0	1	1
Belarus	Russian	224	1	1	1	0	1	1
Belgium	Dutch	159	1	1	1	0	1	1
Bolivia	Spanish	31	1	1	1	1	2	1
Brazil	Portuguese	202	1	1	1	0	1	1
Bulgaria	Bulgarian	141	1	1	1	0	1	1
Cambodia	Central Khmer	7	0	0	0	0	0	0
Canada	English	85	1	1	0	0	1	0
Chile	Spanish	50	1	1	1	1	2	1
Colombia	Spanish	72	1	1	1	1	2	1
Costa Rica	Spanish	5	1	1	1	1	2	1
Croatia	Croatian	246	1	1	1	0	1	1
Cuba	Spanish	29	1	1	1	1	2	1
Cyprus	Greek	16	1	1	1	0	1	1
Czech Republic	Czech	202	1	1	1	0	1	1
Denmark	Danish	96	0	1	1	0	1	0
Dominican Republic	Spanish	22	1	1	1	1	2	1
Ecuador	Spanish	86	1	1	1	1	2	1
Egypt	Arabic	67	1	1	1	1	2	1
El Salvador	Spanish	3	1	1	1	1	2	1
Equatorial Guinea	Spanish	2	1	1	1	1	2	1
Estonia	Estonian	61	0	0	0	0	0	0
Finland	Finnish	213	0	0	0	0	0	0
France	French	592	1	1	1	0	1	1
Germany	German	980	1	0	1	0	1	1
Greece	Greek	104	1	1	1	0	1	1
Guatemala	Spanish	7	1	1	1	1	2	1
Honduras	Spanish	6	1	1	1	1	2	1
India	Hindi	193	1	1	1	0	0	1
Iraq	Arabic	192	1	1	1	1	2	1
Israel	Hebrew	15	1	1	1	1	2	1
Italy	Italian	624	1	1	1	0	1	1
Kazakhstan	Russian	206	1	1	1	0	1	1
Kuwait	Arabic	7	1	1	1	1	2	1
Latvia	Latvian	79	1	1	1	0	1	1
Madagascar	Malagasy	13	0	0	0	0	0	0
Mauritania	Arabic	2	1	1	1	1	2	1
Mexico	Spanish	25	1	1	1	1	2	1
Morocco	Arabic	779	1	1	1	1	2	1
Netherlands	Dutch	186	1	1	1	0	1	1
New Zealand	English	19	1	1	0	0	1	0
Nicaragua	Spanish	3	1	1	1	1	2	1
Niger	Hausa	4	1	1	1	1	2	1
Nigeria	Yoruba	77	0	0	0	0	0	0
Panama	Spanish	3	1	1	1	1	2	1
Peru	Spanish	65	1	1	1	1	2	1
Philippines	English	84	1	1	0	0	1	0
Poland	Polish	814	1	1	1	0	1	1
Portugal	Portuguese	453	1	1	1	0	1	1
Qatar	Arabic	1	1	1	1	1	2	1
Romania	Romanian	428	1	1	1	0	1	1
Russia	Russian	1946	1	1	1	0	1	1
Sao Tome and Principe	Portuguese	7	1	1	1	0	1	1
Saudi Arabia	Arabic	4	1	1	1	1	2	1
Slovakia	Slovak	246	1	0	1	0	1	1
Slovenia	Slovenian	39	1	1	1	1	1	1
Suriname	Dutch	92	1	1	1	0	1	1
Sweden	Swedish	223	0	0	1	0	1	0
Syrian Arab Republic	Arabic	85	1	1	1	1	2	1
Thailand	Thai	47	0	0	0	0	0	0
Tunisia	Arabic	164	1	1	1	1	2	1
Turkey	Turkish	590	0	0	0	0	0	0
United Arab Emirates	Arabic	1	1	1	1	1	2	1
United Kingdom	English	779	1	1	0	0	1	0
United States	English	274	1	1	0	0	1	0
Viet Nam	Vietnamese	49	0	0	0	0	0	0
		$\sum$ 13346						

Table 5: Literature overview

	Data	Sample	Language indicator(s)	Language assigned	Outcome
Gay *et al.* (2013)	World Values Survey (1994-2007)	Native women	All four indicators separately	Dominant language in country of origin;	Participation in the labour market, the credit market, land ownership, & politics
	New Immigrant Survey (NIS)	1st gen immigrant women, 18-65	NG + SB + GA NG + SB + GP	Language spoken at home	Labour force participation
Gay *et al.* (2017)	American Community Survey (2005-2015)	Married 1st gen immigrant women, 25-49, speaking language other than English at home	SB x (GP + GA + NG)	Language spoken at home	Labour force participation
Galor *et al.* (2017)	US Census (2000-) American Community Survey (2000-)	2nd gen immigrant women, older than 24, living with their parents	SB	Language spoken by respondent	Educational attainment
Hechavarría *et al.* (2017)	Global Entrepreneurship Monitor (2001-2005)	Country-level	All four indicators separately	Dominant language	Entrepreneurial activity
Hicks *et al.* (2015)	American Time Use Survey (2003-2012) Current Population Survey (2003-2012)	1st gen immigrants	All four indicators separately SB + GP + GA + NG GA conditional on SB	Dominant language in country of origin	Allocation of household tasks
Jakiela & Ozier (2018)	Afrobarometer (2002 – 2013)	Natives in Kenya, Niger, Nigeria & Uganda	Own indicator	Mother tongue	Female labour force participation & educational attainment
Mavisakalyan (2015)	World Values Survey (1994-2007)	Native women, 18-65, excluding retired and students	GP, three categories	Dominant language in country of origin; Language spoken at home	Labour force participation
Santacreu-Vasut *et al.* (2013)	Quotaproject.org	Country-level	SB + GP + GA + NG	Dominant language	Political quotas
Santacreu-Vasut *et al.* (2014)	World Values Survey (1994-2007)	Country-level	All four indicators separately	Dominant language	Female participation on boards and senior management
	MIX market data		NG + SB + GA NG + SB + GP		
van der Velde *et al.* (2015)	Meta-analysis, studies on the gender wage gap	Country-level	Not specified, presumably SB, GA, GP	Dominant language	Gender wage gap

Table 7: Sample characteristics

	First generation		Second generation	
	mean	s.d.	mean	s.d.
Individual characteristics				
Weekly working hours	37.1	14.41	38.4	12.75
Female (0/1)	0.6	0.50	0.6	0.50
Age in years	40.6	8.31	39.9	8.80
Mother no/primary education (0/1)	0.4	0.49	0.3	0.46
Mother tertiary education (0/1)	0.2	0.38	0.2	0.38
Christian religion (0/1)	0.5	0.50	0.3	0.48
Eastern religion (0/1)	0.0	0.12	0.0	0.09
Islamic religion (0/1)	0.1	0.29	0.1	0.25
Characteristics of the origin country				
FLFP (0-1)	0.4	0.12	0.4	0.14
TFR (number of children born)	2.3	1.05	2.3	1.14
GDP (in 1000 US Dollar)	9.1	9.18	9.6	9.51
Observations	7,399		5,947	