Do Linguistic Structures Affect Human Capital?  
The Case of Pronoun Drop

Horst Feldmann*

I. INTRODUCTION

Recent research in economics suggests that certain linguistic structures affect both individuals’ behavior and aggregate outcomes. For example, Chen (2013) finds languages that grammatically associate the future and the present to foster future-oriented behavior such as thriftiness, and that countries with such languages have substantially higher national savings rates. Galor et al. (2017) report that second-generation migrants in the US who speak languages characterized by the presence of periphrastic future tense (associated with long-term orientation) have a higher probability of attending college. They also find an adverse effect of speaking a language with gender distinctions in grammar on female college attendance. Related papers find gendered languages to be associated with various types of other behavior and outcomes that are detrimental to women, such as more household tasks, lower labor market participation, especially in management roles, and fewer gender quotas in politics.1

This recent research as well as the present paper draw on the linguistic relativity hypothesis. Originating from the linguists Whorf (2012 [1956]) and Sapir (1963), the hypothesis holds that the structure of a language can influence its speakers’ thought and behavior (for a discussion of the hypothesis, see Lucy 1997 and Pütz and Verspoor 2000). By dividing the world into concepts and forcing speakers to encode specific aspects of the world in a particular way,
language shapes speakers’ mental representation of reality. Linguistic categories can thus influence cognitive categories and processes. This can affect the way its speakers think about the world and behave. Over the past few decades, several experimental studies in linguistics and psychology have corroborated the hypothesis (for a survey of these studies, see Everett 2013).

This paper empirically studies the effect of grammatical rules that permit speakers to drop a personal pronoun when it is used as a subject of a sentence (pronoun drop, for short). Davis and Abdurazokzoda (2016) give a good example of such a rule: “pronoun drop is permitted in Spanish, such that the English sentence ‘I speak’ may be translated either as ‘Yo ablo’ or as ‘Ablo,’ dropping the subject pronoun ‘Yo.’ In contrast, pronoun drop is not permitted in English, as the pronoun ‘I’ is required to make sense of the sentence” (p. 544). Our paper studies whether pronoun drop rules affect human capital, specifically educational attainment and educational investment. It uses both individual-level and country-level data, in each case covering a large number of languages and countries. The individual-level analysis finds that speakers of pronoun drop languages have a lower probability of having completed secondary or tertiary education, compared with speakers of non-pronoun drop languages. Consistent with these findings, the country-level analysis establishes that countries where the dominant languages permit pronoun drop have lower secondary school enrollment rates. In both cases, the magnitude of the effect is substantial, particularly among females. The country-level analysis also finds the effect to be direct – i.e., not working through contemporary culture. Furthermore, it finds the effect of pronoun drop languages not to be attenuated by contemporary culture – except very slightly so in the cases of contemporary individualism/collectivism and long-term orientation. Finally, it finds that the various dimensions of contemporary culture do not affect educational investment.

As will be explained in more detail below, our interpretation of these findings is that rules licensing pronoun drop perpetuate ancient cultural values and norms – formed and encoded in those rules in the distant past – that give primacy to the collective over the individual. Through such language rules, these ancestral cultural values and norms could still be effective nowadays – inducing governments and families to invest comparatively little in the education of the young, as education usually increases the independence of the individual from both the state and the extended family and may thus reduce his or her commitment to these institutions. By contrast, modern cultural values and norms have become much more individualistic in recent decades, including in many traditionally collectivist societies, which could explain why we find measures of contemporary culture not to negatively affect education.

The remainder of the paper is organized as follows. Section II explains the background of our empirical exercise and develops a testable hypothesis. While section III presents our individual-level analysis, section IV presents our country-level analysis. Section V concludes.
II. BACKGROUND AND HYPOTHESIS

Darwin (1859) already considered language as a store of information about the pedigree of humankind. Cavalli-Sforza (2000) argues that language was created by culture and acts as its vehicle by encoding aspects of it. Indeed, many modern evolutionary linguists and cognitive scientists believe that today’s linguistic structures are the product of cultural tradition. In their view, such structures have emerged through cultural transmission of language across hundreds or thousands of generations of learners. Thus, contemporary linguistic structures can reflect deep cultural values and norms that became codified in language in the distant past. Once these ancestral cultural traits had been encoded in linguistic structures, they became stable and persistent, reinforcing them and helping them to be transmitted from generation to generation. Thus, linguistic structures sustained ancestral cultural traits over millennia, probably until today (Galor et al. 2017). Because language shapes speakers’ mental representation of reality and because, according to the linguistic relativity hypothesis, these linguistic categories can influence speakers’ thought and behavior, the imprint of ancestral cultural values and norms in linguistic structures implies that these values and norms may still affect speakers’ views and behavior nowadays.

One example of deep and stable features of language that probably reflect distant cultural traditions are the grammatical rules on the use of personal pronouns (Tabellini 2008). Pronoun drop rules allow speakers of the relevant languages to omit the subject of a sentence when it is a personal pronoun. Information about the subject can be recovered either from the verb conjugation, as in Spanish, or from the context of the utterance, as in Japanese, Chinese or Korean (Kashima and Kashima 2003). In other languages the use of pronominal subjects is obligatory. For example, in English “I” or “you” must be included in a sentence even if the referent is unambiguous.

In languages requiring its speakers to mention the subject pronoun explicitly, the person is highlighted as a figure against the context of the speech (Kashima

---

2See, e.g., Christiansen and Kirby (2003), Levinson (2003), Johansson (2005), Kirby et al. (2007), Evans and Levinson (2009), Baronchelli et al. (2015), Thompson et al. (2016).

3Evans (2013) labels the impact of culture on language “Vico-Herder effects”, after two early proponents of the view that languages express unique aspects of the history and worldview of their cultures (or peoples or nations). He cites several examples of such effects, where, according to research by modern evolutionary linguists, some cultural traits ended up shaping the grammar in some way.

4As Hayek (2010 [1942-44], p. 146) put it, “by learning to speak, we learn to classify things in a certain manner without acquiring the actual experiences which have led successive generations to evolve this system of classification.” And “the structure of the language itself implies certain views about the nature of the world; and by learning a particular language we acquire a certain picture of the world, a framework of our thinking within which we henceforth move without being aware of it. As we learn as children to use our language according to rules which we do not explicitly know, so we learn with language not only to act according to the rules of language, but according to many other rules of interpreting the world and of acting appropriately” (Hayek 2014 [1965], p. 44).
and Kashima 1998). This entails a psychological differentiation between the subject and the context, and a focus on the former. For example, if the first-person singular pronoun (e.g., “I” in English) is included in a sentence, the speaker is the focus of attention (Kashima and Kashima 2003). By contrast, dropping a personal pronoun as the subject of a sentence symbolically de-emphasizes the significance of the relevant individual. He is symbolically represented as being embedded in the context. The languages of collectivist cultures do not require the use of ‘I’ (Triandis 2015). In Arabic, a pronoun drop language, there is even the saying, “The satanic term ‘I’ be banned” (Hofstede 2001).

Languages that allow speakers to drop a personal pronoun when used as a subject of a sentence may entrench not only the belief that individuals are embedded in social contexts, but also that they need to submit themselves to the collective. The interests of the collective may be regarded as more important than those of the individual. In the distant past, the custom to drop personal pronouns, especially ‘I’, may have reflected and enshrined cultural traditions that gave more emphasis to the collective – society and the extended family – than the individual. Thus, pronoun drop languages may have perpetuated such collectivist traditions. Conversely, languages that require the use of personal pronouns may instill and reinforce in its speakers the belief, perhaps unconsciously, that individuals and their preferences are important. They may reflect individualistic cultural traditions.

We test the hypothesis that pronoun drop languages negatively affect education, particularly among females. Why may there be such an effect? As explained above, pronoun drop languages may help to entrench the primacy of the collective – the extended family or society at large – over the individual. Under such circumstances, both governments and parents may be inclined to invest relatively little into children’s education (Hofstede 2001, Schwartz 2008). Governments may tend to do so for two reasons in particular. First, they may view the education of children primarily as the continued responsibility of the extended family. Second, they may want to avoid as much as possible that well-educated individuals become independent of the state and, in particular, that they emigrate. Families may want to avoid as much as possible that well-educated children move to another town or country because this would loosen family ties and obligations such as caring for elderly members of the family. These views of governments and family are likely to be particularly strong with respect to girls. In collectivist societies, their education is generally seen as less important than the education of boys. Additionally, the advancement of girls is usually seen as more of a threat to the extended family because females are expected to bear the bulk of the burden when it comes to caring for elderly family members. Lastly, in countries dominated by pronoun drop languages both males and females may be more willing to conform with these views of family and government because the pronoun drop rules may have instilled in them from an early age the subconscious notion that they have to do so.
By contrast, the focus on the individual in non-pronoun drop languages may entail comparatively more educational investment (Hofstede 2001, Schwartz 2008). In societies dominated by such languages, there may be the general belief – including among parents and policy makers – that the education of children is desirable in order to develop their personalities, their unique interests and abilities. Their educational advancement may thus be seen positively, not as a threat to the family or the state. This attitude is likely to pertain to the education of girls too. Furthermore, languages that forbid dropping the first-person pronoun may contribute to instill in both boys and girls from an early age a sense of their uniqueness, nurturing educational aspirations as they get older. Thus, the focus on the individual in non-pronoun drop languages may lead to higher school enrollment rates and higher levels of educational attainment, including among females.

III. INDIVIDUAL-LEVEL ANALYSIS

III.1. Data, variables and methodology

The data for our individual-level analysis comes from the World Values Survey, nationally representative surveys of people’s values and beliefs, which also collects information about participants’ characteristics. Since wave 3 (1994-99), the World Values Survey includes a question about the language normally spoken at home. Thus our sample uses data from this wave as well as from waves 4 (1999-2004), 5 (2005-09) and 6 (2010-14). The main source of linguistic information is the World Atlas of Language Structures Online. This source provides the most authoritative information on a large number of languages. For example, it documents for each language covered all grammatical rules on the expression of pronominal subjects (Dryer 2013). Using this information, we created our variable of interest, ‘pronoun drop language’, a dummy that equals 1 if a respondent normally speaks at home a language that permits its speakers to drop a personal pronoun when used as a subject of a sentence. Languages that do not allow pronoun drop were coded 0. We were able to code a total of 103 languages, 86 of which allow pronoun drop. In our individual-level sample, which includes data on 114,894 individuals from 75 countries, 73% of participants speak such a language (for a list of countries included in our individual-level analysis, see Table A1 in the Supporting Information file on the publisher’s article web page).

The World Values Survey also includes information about participants’ highest educational level attained. We used this information to create our dependent variable, ‘secondary or tertiary education’. It is a dummy that takes the value 1 if the highest level attained was either completed secondary or tertiary education. In our sample, 62% of respondents are in this bracket. We control for numerous demographic characteristics (for definitions and descriptive statistics of all variables used
in our individual-level analysis, see Table A2 in the online Supporting Information file). Specifically, we control for ethnicity, age (including age squared), health, marital status and number of children. Furthermore, we control for employment status, income, social class, religion adherence and the size of a respondent’s residential location. Finally, we control for unobserved country and year effects throughout.

A downside of the information on education provided in the World Values Survey, and thus of the dependent variable in our individual-level analysis, is that it does not pertain to contemporaneous schooling but rather to schooling that took place several years, and in many cases several decades, earlier. However, this should hardly be a problem when estimating the effect of ‘pronoun drop language’ on education because the language normally spoken at home is likely to have changed in very few cases over the years. Most people continue to normally speak at home their mother tongue, i.e., the language acquired well before they started going to school.

By contrast, the values of many control variables will have changed since leaving school. Indeed, the highest educational level attained many years ago is likely to affect current income, employment and marital status, in particular. Thus for some controls, there is likely to be a reverse causality bias. For this reason, we omit in our first specification variables for which such a bias is most likely. This specification only includes variables that either clearly are not subject to reverse causality, such as ethnicity, or are little likely to suffer from such a bias, such as religion adherence.

In our second specification, we use all controls, including those that are highly likely to suffer from reverse causality bias. We do so for two reasons. First, these demographic controls are commonly employed in individual-level regressions. Second, in our case one can regard some of them that are most likely to suffer from reverse causality bias as proxies for determinants of participants’ education. For example, as incomes are persistent across generations in most countries (e.g., Jäntti and Jenkins 2014), one can view a participant’s current income as a proxy for his parents’ income, which in turn probably has been one factor determining the education he received (e.g., Taubman 1989, Liu 2003).

Using probit, we estimate the following model:

\[ E_{j,i,t} = \eta + \theta L_{j,i,t} + \sum_{k=1}^{r} \mu_k Z_{k,j,i,t} + \phi_i + \psi_t + \omega_{j,i,t} \]  \hspace{1cm} (1)

\( E_{j,i,t} \) denotes our dummy for completed secondary or tertiary education as the highest educational level attained by individual \( j \) in country \( i \) and survey year \( t \). \( L_{j,i,t} \) is our variable of interest, ‘pronoun drop language’, and \( Z_{k,j,i,t} \) denotes a vector of \( r \) control variables. While \( \eta \) is the constant, \( \phi_i \) and \( \psi_t \) denote country and year dummies, respectively. Finally, \( \omega_{j,i,t} \) represents the error term.

5None of our control variables is highly correlated with ‘pronoun drop language’.
III.2. Results and discussion

Table 1 reports conditional marginal effects of probit regressions, calculated at the means. While columns 1 and 2 present the results for women, columns 3 and 4 present the results for men. Whereas regressions 1 and 3 use the core controls only, regressions 2 and 4 use all of them. Table 1 also reports two goodness-of-fit measures. The first is the most commonly used Pseudo $R^2$ constructed by McFadden (1973). As this measure has been found to have a downward bias (e.g., Veall and Zimmermann 1996), we additionally report the Pseudo $R^2$ constructed by McKelvey and Zavoina (1975). The latter is often regarded as the best fit measure for limited dependent variable models (e.g., Veall and Zimmermann 1996, Long 1997). It has also the advantage of being most comparable to $R^2$ from OLS regressions (e.g., Veall and Zimmermann 1996). Anyhow, in our case both McFadden’s as well as McKelvey and Zavoina’s Pseudo $R^2$ suggest that our model has a very good fit, particularly when including all controls (Table 1).

In all four individual-level regressions, the coefficient on our variable of interest, ‘pronoun drop language’, is negative and significant, mostly at the 1% level. Additionally, we find that the magnitude of the effect is substantial and slightly larger for women. Specifically, women who speak a pronoun drop language are 9-11 percentage points less likely to have completed secondary or tertiary education than women who speak a non-pronoun drop language. For men, the probability is 8-10 percentage points. Thus, the individual-level estimates corroborate our hypothesis.

A brief comment on the results for the controls. Consistent with the literature (e.g., Norton and Tomal 2009, Cooray and Potrafke 2011), we find Islam to have a negative effect on education, particularly among women. Furthermore, we find that both Eastern Orthodoxy and Hinduism may also negatively affect women’s education, while Buddhism may negatively affect the education of men. These results, too, are in line with some other studies (e.g., Norton and Tomal 2009). Also in line with the previous literature (e.g., Bertinelli and Zou 2008), we find that living in a larger town increases the probability of having attained completed secondary or tertiary education. Lastly and unsurprisingly, belonging to a higher social class increases the probability of having completed secondary or tertiary education as well.

IV. COUNTRY-LEVEL ANALYSIS

IV.1. Variables, sample and methodology

Here, our variable of interest is ‘pronoun drop’, measuring the decimal fraction of a country’s population speaking a language that permits to drop a
| Dependent variable: | (1) Females | (2) Females | (3) Males | (4) Males |
|---------------------|-------------|-------------|-----------|-----------|
|                     | Core controls | All controls | Core controls | All controls |
| Pronoun drop language | -0.107*** (0.041) | -0.087*** (0.031) | -0.102** (0.040) | -0.081*** (0.030) |
| Town                | 0.024*** (0.003) | 0.019*** (0.003) | 0.026*** (0.002) | 0.022*** (0.003) |
| White               | -0.015 (0.024) | -0.013 (0.025) | -0.004 (0.023) | -0.012 (0.024) |
| Social class        | 0.129**** (0.007) | 0.088**** (0.007) | 0.131**** (0.006) | 0.094**** (0.005) |
| Protestant          | -0.017 (0.020) | -0.006 (0.019) | 0.001 (0.014) | -0.003 (0.014) |
| Roman Catholic      | -0.026* (0.013) | -0.021 (0.013) | -0.019 (0.014) | -0.019 (0.014) |
| Eastern Orthodox    | -0.047** (0.020) | -0.041** (0.018) | -0.030 (0.022) | -0.029 (0.022) |
| Other Christian     | 0.003 (0.025) | 0.020 (0.021) | 0.014 (0.024) | 0.014 (0.024) |
| Jewish              | -0.028 (0.076) | -0.016 (0.062) | 0.011 (0.045) | 0.028 (0.052) |
| Muslim              | -0.159**** (0.031) | -0.116**** (0.027) | -0.119**** (0.027) | -0.099**** (0.025) |
| Hindu               | -0.117**** (0.030) | -0.083**** (0.024) | -0.034 (0.055) | -0.030 (0.058) |
| Buddhist            | -0.052** (0.023) | -0.042 (0.026) | -0.098**** (0.032) | -0.090**** (0.033) |
| Other Eastern religions | -0.073 (0.081) | -0.047 (0.080) | -0.130 (0.111) | -0.081 (0.100) |
| Other religions     | -0.022 (0.015) | -0.004 (0.019) | -0.006 (0.016) | 0.009 (0.020) |
| Age                 | 0.003 (0.002) | 0.005*** (0.002) | 0.007*** (0.002) | 0.009*** (0.002) |
| Age$^2$             | -0.000*** (0.000) | -0.000*** (0.000) | -0.000*** (0.000) | -0.000*** (0.000) |
| Income              | 0.028*** (0.002) | 0.030*** (0.002) | 0.030*** (0.002) | 0.030*** (0.002) |
| Employed            | 0.136*** (0.015) | 0.025* (0.013) | 0.025* (0.013) | 0.025* (0.013) |
| Unemployed          | 0.042*** (0.016) | -0.041** (0.016) | -0.041** (0.016) | -0.041** (0.016) |
| Health              | 0.035*** (0.005) | 0.027*** (0.004) | 0.027*** (0.004) | 0.027*** (0.004) |
| Married             | -0.024* (0.013) | -0.002 (0.011) | -0.002 (0.011) | -0.002 (0.011) |
| Living together     | -0.085*** (0.014) | -0.045*** (0.014) | -0.045*** (0.014) | -0.045*** (0.014) |
| Divorced            | 0.007 (0.019) | 0.026 (0.020) | 0.026 (0.020) | 0.026 (0.020) |
| Separated           | -0.025 (0.018) | -0.039** (0.016) | -0.039** (0.016) | -0.039** (0.016) |

(Continues)
personal pronoun when it is used as the subject of a sentence (for definitions, descriptive statistics and sources of all country-level variables, see Table A3 in the online Supporting Information file). For this variable, our data is mainly from Davis and Abdurazokzoda’s (2016) recently constructed linguistic dataset, which overcomes weaknesses of Kashima and Kashima’s (1998) original data (for a discussion of these weaknesses, see Davis and Abdurazokzoda 2016). Using the same sources and methodology as Davis and Abdurazokzoda (2016), we coded the ‘pronoun drop’ variable for several more countries. The country-level measure of pronoun drop has been constructed by taking into account, for each country, up to three popularly spoken languages.

Our educational investment variables refer to secondary schooling. Specifically, we use secondary enrollment rates, which are defined as children enrolled in secondary education, regardless of age, as a percentage of children in the age group that officially corresponds to this level of education. We do not use tertiary enrollment rates because in almost all developing countries these rates have been very low over our sample period. Also, data availability for tertiary enrollment rates is much more limited.

As in our individual-level analysis, we estimate regressions separately for females and males. As explained in section II, countries dominated by pronoun drop languages are likely to restrict the schooling of girls in particular. Our separate regressions are intended to test this part of our hypothesis.
We use a large number of variables to control for the impact of other potential determinants of education.\textsuperscript{6} The control variables we employ have been selected on the basis of the relevant theoretical and empirical literature. For brevity, instead of surveying this literature in detail let us just list the variables and cite some of the papers that have found the respective variable to be potentially important. To start with, we control for public spending on education (e.g., Trostel 2002). Moreover, we control for political rights and civil liberties (e.g., Lake and Baum 2001). We employ several demographic variables: life expectancy (e.g., Soares 2005), urbanization rate (e.g., Bertinelli and Zou 2008) and, in some robustness checks, population growth rate (e.g., Becker and Lewis, 1973), death rate (e.g., Forston 2011) and the shares of children and elderly in the population (e.g., Poterba 1997). Throughout, we control for major religions (e.g., Feldmann 2016a). We also control for relevant economic characteristics. Specifically, we use GDP per capita (e.g., Mincer 1996), GDP growth rate as a proxy for business cycle fluctuations (e.g., Méndez and Sepúlveda 2012), private credit as a proxy for credit constraints (e.g., De Gregorio 1996), openness (e.g., Baskaran and Hessami 2012) and, in one robustness check, economic freedom (e.g., Feldmann 2017). Furthermore, we control for geographic conditions (e.g., Gallup et al. 1999) and, in one robustness check, regional characteristics. As educational investments are future-oriented, we control for languages with strong future-time reference in a further robustness check (e.g., Chen 2013). While in one robustness check we add controls for linguistic, ethnic and religious fractionalization (e.g., Alesina et al. 2003), in another one we additionally control for colonial history (e.g., Feldmann 2016b).

As explained previously, pronoun drop rules may reflect ancient values and norms of collectivism. As cultural values and norms usually change slowly, including corresponding measures of contemporary culture in our regressions could attenuate or nullify any direct effect of ‘pronoun drop’. To shed light on this issue, we include in some additional regressions, one at a time, survey-based indicators of relevant aspects of contemporary culture. We start with ‘embeddedness’, ‘autonomy’ and ‘individualism’. While ‘embeddedness’ measures the extent to which a culture views people as entities embedded in the collectivity, ‘autonomy’ measures the extent to which it encourages individuals to independently pursue their ideas and experiences. ‘Individualism’ measures the extent to which people are supposed to look after themselves and their immediate family only, as opposed to being strongly integrated and loyal to a cohesive group such as the extended family (collectivism). Higher values of this measure indicate a higher level of individualism, whereas lower values indicate a higher level of collectivism.

\textsuperscript{6}None of our control variables is highly correlated with ‘pronoun drop’.
In some additional regressions, we alternatively include a survey-based measure of contemporary long-term orientation. This is relevant here because, as indicated above, the rewards of most forms of human capital investment materialize in the long term. For example, the return to schooling accrues throughout working life. The variable ‘long-term orientation’ measures the extent to which society fosters pragmatic virtues oriented towards future rewards.

In our country-level analysis, our sample consists of 101 countries (Table A4 in the online Supporting Information file). In 64 of them, the most widely spoken languages permit pronoun drop. For three reasons, we analyze country-level in addition to individual-level data. First, it enables us to control for contemporary culture as well as for numerous other national characteristics. Second, it also enables us to estimate the effect on national rates of schooling. Such rates are important for economic and social development. Third, related research suggests that countries are meaningful units of analysis. For example, Schwartz (2004) demonstrates that cultural values are much more similar within than between countries, differences between ethnic groups and regions within countries notwithstanding. Similarly, Inglehart and Baker (2000) report that the historically dominant religion of a country shapes all inhabitants – including those adhering to a different religion, or no religion at all – into a given national culture. Perhaps a country’s dominant languages also shape all (or most) of its inhabitants’ values and norms – even if some have a different mother tongue.

As our variable of interest, ‘pronoun drop’, is time-invariant, we use cross-country rather than panel data. The data for the time-variant variables has been averaged over the period 1972-2012. As with our individual-level analysis, the size of our sample is determined by data availability only, both with respect to countries and with respect to the time period. Using data from large samples should lead to general results.

We estimate the following OLS model:

\[ S_i = \alpha + \beta P_i + \sum_{j=1}^{q} \gamma_j X_{j,i} + \epsilon_i \]  

\( S_i \) is a secondary enrollment rate variable of country \( i \), covering either girls or boys. \( P_i \) denotes the ‘pronoun drop’ variable and \( X_{j,i} \) a vector of \( q \) control variables. While \( \alpha \) is the constant term, \( \epsilon_i \) represents the error term.

We do not instrument for ‘pronoun drop’ because this variable, as the variable ‘pronoun drop language’ in our individual-level analysis, is most likely to be
exogenous. Linguistic structures such as rules governing the use of personal pronouns have been formed in the distant past and evolve very slowly, over generations.

By contrast, most of our control variables are probably endogenous. For several reasons, we do not instrument for them either. First, instrumenting for several variables at once leads to complicated problems of identification. Second, for many of our potentially endogenous controls, there are no plausible instruments. Third, several variables that have been used as instruments in previous papers, such as geographic variables, are likely to directly affect education. Fourth, many of these potential instruments are not specific enough to the variables we would like to instrument for. Excluded instruments need to be specific to the instrumented variable (Acemoglu 2005). Finally, most of the potential instruments, such as geographic variables or legal origins, have been used in previous studies as instruments for various other variables. A variable can be a valid instrument in at most one such study (Bazzi and Clemens 2013). As, for these reasons, we do not instrument for any of our controls, our estimates of most of these variables are subject to endogeneity bias.

However, our estimates for ‘pronoun drop’ are unlikely to have such a problem. Endogeneity bias has three main dimensions: reverse causality bias, omitted variable bias and measurement error. As explained above, our variable of interest is most unlikely to be subject to the first. As we control for all major factors that have been found to determine education in large samples of countries, it is also unlikely to suffer from the second. And the third dimension should not be an issue either, given the high-quality data and methodology we use to measure ‘pronoun drop’.

IV.2. Results and discussion

Tables 2–5 present the regression results using country-level data. While Table 2 reports the results from our baseline specification, Tables 3 and 4 report the results from our robustness checks – Table 3 for girls and Table 4 for boys. Table 5 presents the results from regressions in which we add contemporary culture variables – columns 1-5 for girls and columns 6-10 for boys. For brevity, the estimates for the main control variables are omitted in Tables 3–5. Each of the regressions in these tables additionally uses the same controls as the baseline regressions of Table 2. Note that our country-level regressions explain about 80% of the variability in the data. Thus, the overall fit of the equations is very good.

The estimates for our variable of interest, ‘pronoun drop’, are consistent with our individual-level estimates. In all country-level regressions, the coefficient on ‘pronoun drop’ is statistically significant, mostly at the 1% level. Its algebraic sign is negative, suggesting that during the sample period, pronoun drop
languages had a detrimental impact on secondary school enrollment. Also in line with our individual-level estimates, we find the magnitude of the effect to be substantial, particularly among girls. Specifically, countries in which the popularly spoken languages permit personal pronoun drop had a secondary enrollment rate that was approximately 10-11 percentage points lower among girls and about 9-10 percentage points lower among boys, compared with countries in which the popularly spoken languages require the use of personal pronouns (Tables 2–5). Thus, the estimates corroborate our hypothesis, including that the

Table 2

Country-level main estimates

| Variable                        | (1) Female secondary enrollment rate | (2) Male secondary enrollment rate |
|---------------------------------|--------------------------------------|------------------------------------|
| Pronoun drop                    | -11.50***                            | -10.53***                          |
|                                 | (3.35)                               | (3.03)                             |
| Public spending on education    | -40.02                               | -35.03                             |
|                                 | (68.78)                              | (65.87)                            |
| Political rights & civil liberties | 5.49                                 | -0.82                              |
|                                 | (9.97)                               | (9.57)                             |
| Life expectancy                 | 1.78***                              | 1.51***                            |
|                                 | (0.37)                               | (0.36)                             |
| Urbanization rate               | 20.39**                             | 8.29                               |
|                                 | (8.94)                               | (8.72)                             |
| Christian population share      | -9.82                                | -8.91                              |
|                                 | (9.59)                               | (8.99)                             |
| Muslim population share         | -18.44****                           | -10.63                             |
|                                 | (6.81)                               | (6.95)                             |
| Eastern religions population share | -9.59                               | -1.98                              |
|                                 | (10.32)                              | (10.45)                            |
| GDP per capita                  | -3.05                                | 1.47                               |
|                                 | (3.77)                               | (3.59)                             |
| GDP growth rate                 | -3.93***                             | -3.63***                           |
|                                 | (1.24)                               | (1.09)                             |
| Private credit                  | 0.81                                 | -2.19                              |
|                                 | (6.12)                               | (5.31)                             |
| Openness                        | 12.80**                             | 7.56                               |
|                                 | (5.98)                               | (5.80)                             |
| Tropical area                   | -11.59**                             | -14.78***                          |
|                                 | (4.80)                               | (4.62)                             |
| Navigable waters                | -4.63                                | -4.30                              |
|                                 | (6.71)                               | (6.19)                             |
| Number of observations          | 101                                  | 101                                |
| $R^2$                           | 0.82                                 | 0.79                               |
| $F$ statistic                   | 61.04***                             | 36.46***                           |
| Root mean squared error         | 12.83                                | 12.60                              |

Notes: OLS estimation. Dependent variables: ‘female secondary enrollment rate’ (column 1) and ‘male secondary enrollment rate’ (column 2). The data for both ‘pronoun drop’ and the geographic variables are time-invariant. The data for all other variables have been averaged over 1972-2012. Both regressions additionally include a constant term. Robust standard errors are reported in parentheses. ***(***) denotes statistically significant at the 1%(5%) level.
### Table 3

Country-level robustness checks females

| Variable                              | (1) Strong future-time reference added | (2) Fractionalization variables added | (3) Colonial variables added | (4) Population growth rate added | (5) Child and elderly population shares added | (6) Death rate added | (7) Economic freedom added | (8) Regional dummies added |
|---------------------------------------|----------------------------------------|--------------------------------------|------------------------------|----------------------------------|---------------------------------------------|----------------------|--------------------------|--------------------------|
| Pronoun drop                          | -12.58***                              | -8.03**                              | -7.61**                      | -10.33***                        | -9.99***                                    | -11.49***            | -9.14***                  | -9.71***                 |
|                                       | (3.62)                                 | (3.47)                               | (3.60)                       | (3.61)                           | (3.46)                                      | (3.40)                | (2.83)                   | (3.34)                   |
| Strong future-time reference          | 4.21                                   |                                      |                              |                                  |                                             |                      |                          |                          |
|                                       | (3.87)                                 |                                      |                              |                                  |                                             |                      |                          |                          |
| Linguistic fractionalization          |                                        | -2.02                                |                              |                                  |                                             |                      |                          |                          |
|                                       |                                        | (11.84)                              |                              |                                  |                                             |                      |                          |                          |
| Ethnic fractionalization              |                                        | -10.97                               |                              |                                  |                                             |                      |                          |                          |
|                                       |                                        | (13.42)                              |                              |                                  |                                             |                      |                          |                          |
| Religious fractionalization           |                                        | 13.26                                |                              |                                  |                                             |                      |                          |                          |
|                                       |                                        | (8.95)                               |                              |                                  |                                             |                      |                          |                          |
| Former British colony                 |                                        |                                      |                              |                                  |                                             |                      |                          |                          |
|                                       |                                        | 0.80                                 |                              |                                  |                                             |                      |                          |                          |
|                                       |                                        | (4.12)                               |                              |                                  |                                             |                      |                          |                          |
| Former Spanish colony                 |                                        | -14.45*                              |                              |                                  |                                             |                      |                          |                          |
|                                       |                                        | (7.94)                               |                              |                                  |                                             |                      |                          |                          |
| Former French colony                  |                                        | -12.60***                            |                              |                                  |                                             |                      |                          |                          |
|                                       |                                        | (3.91)                               |                              |                                  |                                             |                      |                          |                          |
| Population growth rate                |                                        |                                      |                              |                                  |                                             | -3.07                |                          |                          |
| Child population share                |                                        |                                      |                              |                                  |                                             |                      | -1.34**                  |                          |
|                                       |                                        |                                      |                              |                                  |                                             |                      | (0.59)                   |                          |
| Elderly population share              |                                        |                                      |                              |                                  |                                             |                      | -1.59*                   |                          |
|                                       |                                        |                                      |                              |                                  |                                             |                      | (0.94)                   |                          |
| Death rate                            |                                        |                                      |                              |                                  |                                             | 1.87                 |                          |                          |
| Economic freedom                      |                                        |                                      |                              |                                  |                                             |                      | 58.87*                   |                          |
|                                       |                                        |                                      |                              |                                  |                                             |                      | (30.40)                  |                          |

(Continues)
Table 3. (Continued)

| (1) Strong future-time reference added | (2) Fractionalization variables added | (3) Colonial variables added | (4) Population growth rate added | (5) Child and elderly population shares added | (6) Death rate added | (7) Economic freedom added | (8) Regional dummies added |
|----------------------------------------|--------------------------------------|-------------------------------|----------------------------------|-----------------------------------------------|---------------------|---------------------------|--------------------------|
| America                                | -25.58                               | (18.98)                       | -18.81                          | (19.64)                                      | -21.71                           | (19.15)                   | -15.32                   | (20.45)                   |
| Africa & the Middle East               | -12.56                               | (19.84)                       | 10.52                           | (18.94)                                      | 12.78                            | (20.34)                   | 12.56                    | (20.34)                   |
| Europe                                 | -15.26                               | (20.15)                       | 10.52                           | (18.94)                                      | 12.78                            | (20.34)                   | 12.56                    | (20.34)                   |
| Asia                                   | -12.56                               | (19.84)                       | 10.52                           | (18.94)                                      | 12.78                            | (20.34)                   | 12.56                    | (20.34)                   |

Main control variables: Yes

Number of observations: 93 96 101 101 99 101 86 101

$R^2$: 0.80 0.84 0.84 0.82 0.84 0.82 0.84 0.84

$F$ statistic: 38.27*** 53.81*** 62.82*** 60.46*** 57.01*** 57.30*** 43.74*** 46.71***

Root mean squared error: 12.56 12.45 12.14 12.78 12.45 12.90 12.31 12.52

Notes: OLS estimation. Dependent variable: "female secondary enrollment rate". In addition to the explanatory variables mentioned above, each regression uses the same control variables as the main regressions reported in Table 2. For brevity, the estimates for these main control variables are omitted. The data for the linguistic, fractionalization, colonial, regional and geographic variables are time-invariant. The data for all other variables have been averaged over 1972-2012. Each regression also includes a constant term. Robust standard errors are reported in parentheses. ***(**/) denotes statistically significant at the 1%(5%/10%) level.
Table 4
Country-level robustness checks males

|                                | (1)                      | (2)                      | (3)                      | (4)                      | (5)                      | (6)                      | (7)                      | (8)                      |
|--------------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
|                                | Strong future-time       | Fractionalization        | Colonial                  | Population growth rate   | Child and elderly        | Death rate               | Economic freedom         | Regional dummies         |
|                                | reference added          | variables added          | variables added           | rate added               | population shares added  | added                    | added                    | added                    |
| Pronoun drop                   | -11.23***                | -7.43**                  | -7.03**                  | -9.40***                 | -9.34***                 | -10.52***                | -8.38***                 | -8.90***                 |
|                                | (3.50)                   | (3.17)                   | (3.19)                   | (3.20)                   | (3.18)                   | (3.07)                   | (2.54)                   | (2.93)                   |
| Strong future-time reference   | 3.32                     | 0.46                     | 14.70                    | 14.70                    | 14.70                    | 14.70                    | 14.70                    | 14.70                    |
|                                | (3.78)                   | (11.02)                  | (9.33)                   | (13.15)                  | (9.33)                   | (13.15)                  | (9.33)                   | (13.15)                  |
| Linguistic fractionalization   |                          |                          |                          |                          |                          |                          |                          |                          |
|                                |                          |                          |                          |                          |                          |                          |                          |                          |
| Ethnic fractionalization       | -11.76                   | 0.25                     |                          |                          |                          |                          |                          |                          |
|                                | (13.15)                  | (3.93)                   |                          |                          |                          |                          |                          |                          |
| Religious fractionalization    |                          |                          |                          |                          |                          |                          |                          |                          |
|                                |                          |                          |                          |                          |                          |                          |                          |                          |
| Former British colony          |                          |                          |                          |                          |                          |                          |                          |                          |
|                                |                          |                          |                          |                          |                          |                          |                          |                          |
| Former Spanish colony          | -13.38*                  |                          |                          |                          |                          |                          |                          |                          |
|                                | (7.66)                   |                          |                          |                          |                          |                          |                          |                          |
| Former French colony           | -13.30***                |                          |                          |                          |                          |                          |                          |                          |
|                                | (3.54)                   |                          |                          |                          |                          |                          |                          |                          |
| Population growth rate         |                          |                          |                          |                          |                          | -2.98                    |                          |                          |
|                                |                          |                          |                          |                          |                          | (2.16)                   |                          |                          |
| Child population share         |                          |                          |                          |                          | -1.41**                  |                          |                          |                          |
|                                |                          |                          |                          |                          | (0.57)                   |                          |                          |                          |
| Elderly population share       |                          | -1.56*                   |                          |                          |                          |                          |                          |                          |
|                                |                          | (0.85)                   |                          |                          |                          |                          |                          |                          |
| Death rate                     |                          |                          |                          |                          |                          | 2.03                     |                          |                          |
|                                |                          |                          |                          |                          |                          | (6.95)                   |                          |                          |
| Economic freedom               |                          |                          |                          |                          |                          |                          | 55.06*                   |                          |
|                                |                          |                          |                          |                          |                          |                          | (29.83)                  |                          |

(Continues)
| Main control variables | (1) Strong future-time reference added | (2) Fractionalization variables added | (3) Colonial variables added | (4) Population growth rate added | (5) Child and elderly population shares added | (6) Death rate added | (7) Economic freedom added | (8) Regional dummies added |
|------------------------|---------------------------------------|-------------------------------------|--------------------------|---------------------------------|---------------------------------|-----------------|------------------------|------------------------|
| America                | Yes                                   | Yes                                 | Yes                      | Yes                             | Yes                             | Yes             | Yes                     | -28.34                 |
|                        |                                       |                                     |                          |                                 |                                 |                 |                        | (20.00)                |
| Africa & the Middle East|                                      |                                      |                          |                                 |                                 |                 |                        | -21.28                 |
|                        |                                       |                                     |                          |                                 |                                 |                 |                        | (20.45)                |
| Europe                 |                                      |                                      |                          |                                 |                                 |                 |                        | -22.73                 |
|                        |                                       |                                     |                          |                                 |                                 |                 |                        | (20.07)                |
| Asia                   |                                      |                                      |                          |                                 |                                 |                 |                        | -18.85                 |
|                        |                                       |                                     |                          |                                 |                                 |                 |                        | (21.16)                |
| Number of observations | 93                                    | 96                                  | 101                      | 101                             | 99                              | 101             | 86                      | 101                    |
| $R^2$                  | 0.76                                  | 0.81                                | 0.82                     | 0.79                            | 0.81                            | 0.79            | 0.81                    | 0.81                   |
| $F$ statistic          | 23.59***                             | 30.21***                            | 49.74***                 | 33.70***                        | 32.87***                        | 33.42***        | 28.61***                | 27.76***               |
| Root mean squared error| 12.59                                | 12.36                               | 11.96                    | 12.55                           | 12.18                           | 12.67           | 12.03                   | 12.17                  |

Notes: OLS estimation. Dependent variable: ‘male secondary enrollment rate’. In addition to the explanatory variables mentioned above, each regression uses the same control variables as the main regressions reported in Table 2. For brevity, the estimates for these main control variables are omitted. The data for the linguistic, fractionalization, colonial, regional and geographic variables are time-invariant. The data for all other variables have been averaged over 1972-2012. Each regression also includes a constant term. Robust standard errors are reported in parentheses. ***(**/***) denotes statistically significant at the 1%(5%/10%) level.
Table 5

Culture variables added in country-level regressions

|                  | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
|                  | Female secondary enrollment rate | Male secondary enrollment rate |
| Embeddedness     |     |     |     |     |     |     |     |     |     |      |
| Added            | Embeddedness added | Autonomy added | Individualism added | Long-term orientation added | Long-term orientation and interaction added |
| Pronoun drop     | -11.26*** | -11.11*** | -11.75** | -11.78*** | -31.85** | -9.78*** | -10.16*** | -9.09* | -10.97*** | -30.09*** |
|                  | (3.36) | (3.68) | (4.96) | (3.97) | (12.76) | (2.92) | (3.40) | (4.52) | (3.25) | (9.80) |
| Embeddedness     | 11.35 |     |     |     |     | 9.78 |     |     |     |      |
|                  | (12.30) |     |     |     |     | (12.82) |     |     |     |      |
| Autonomy         |     |     | -1.96 |     | -1.96 | -6.19 |     |     |     |      |
|                  |     |     | (10.07) |     | (10.07) | (10.01) |     |     |     |      |
| Individualism    |     |     | 20.69 |     | 20.69 | 26.37 |     |     |     |      |
|                  |     |     | (19.21) |     | (19.21) | (19.81) |     |     |     |      |
| Long-term orientation |     |     |     | -4.40 | -25.98* |       |     |     |     |      |
|                  |     |     |     | (9.64) | (15.14) |     |     |     |     |      |
| Pronoun drop x   |     |     |     | 41.80* |       |     |     |     |     |      |
| Long-term orientation |     |     |     |     | (20.94) |     |     |     |     |      |
| Main control variables | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of observations | 62 | 62 | 61 | 61 | 73 | 62 | 62 | 62 | 73 | 73 |
| $R^2$            | 0.87 | 0.87 | 0.83 | 0.82 | 0.84 | 0.83 | 0.83 | 0.80 | 0.79 | 0.82 |
| $F$ statistic    | 47.52*** | 45.06*** | 26.47*** | 35.09*** | 41.07*** | 25.81*** | 25.49*** | 16.87*** | 21.88*** | 23.88*** |
| Root mean squared error | 11.43 | 11.60 | 12.29 | 13.55 | 12.71 | 11.66 | 11.73 | 12.18 | 12.91 | 12.31 |

Notes: OLS estimation. Dependent variables: ‘female secondary enrollment rate’ (columns 1-5) and ‘male secondary enrollment rate’ (columns 6-10). In addition to the explanatory variables mentioned above, each regression uses the same control variables as the main regressions reported in Table 2. For brevity, the estimates for these main control variables are omitted. The data for ‘pronoun drop’ as well as the geographic and contemporary culture variables are time-invariant. The data for all other variables have been averaged over 1972-2012. Each regression also includes a constant term. Robust standard errors are reported in parentheses. ***(**/*) denotes statistically significant at the 1%(5%/10%) level.
magnitude of the negative effect is likely to be larger among girls (section II). However, as in the case of our individual-level regressions the difference in the estimated effects between girls and boys is small.

As indicated above, we also explore the effect of pronoun drop languages in the context of contemporary culture. With respect to the latter, we focus on those measures that are most likely to be relevant here: ‘embeddedness’, ‘autonomy’, ‘individualism’ and ‘long-term orientation’. In a first exercise, we add these variables to our baseline specification, one at a time. As reported in Table 5, columns 1-4 and 6-9, each of them is statistically insignificant. By contrast, ‘pronoun drop’ remains statistically significant, in most cases at the 1% level. However, when adding ‘individualism’ the significance of ‘pronoun drop’ falls to 5% in the female regression and to just 10% in the male regression. Also, in the latter regression the absolute size of the coefficient on ‘pronoun drop’ falls substantially. In this context, it is also noteworthy that the absolute size of the correlation coefficient between ‘pronoun drop’, on the one hand, and measures of contemporary culture, on the other, is the highest for ‘individualism’ (Table A5 in the online Supporting Information file).

In a second exercise with the contemporary culture variables, we additionally interact them with ‘pronoun drop’. The intention here is to check whether the magnitude of the negative effect of the latter varies with the degree of one or more of the four dimensions of contemporary culture. With respect to ‘embeddedness’, ‘autonomy’ and ‘individualism’, we find no such interaction effect (results not reported here). However, we find that the negative effect of ‘pronoun drop’ is weaker in countries with more long-term orientation (columns 5 and 10 of Table 5). In the case of girls, this interaction effect is only marginally significant though.

To explore the possibility that ‘pronoun drop’ may mask any effect of contemporary culture, we exclude the former from regressions 1-4 and 6-9 of Table 5 in a third exercise, finding that each of the contemporary measures remains statistically insignificant. Since these regressions obviously suffer from omitted variable bias, we do not report them here.

In a fourth exercise, we explore whether pronoun drop languages affect secondary schooling through contemporary culture. Specifically, we now use ‘pronoun drop’ as an instrument for each of the four contemporary measures, one at a time. This exercise closely resembles several recent papers that use Kashima and Kashima’s (1998) pronoun drop data to instrument for a variety

---

8For completeness, we additionally perform all contemporary culture exercises with all other indices constructed by Schwartz (2004, 2006) and Hofstede (2014). The other Schwartz indices are ‘hierarchy’, ‘egalitarianism’, ‘mastery’ and ‘harmony’. The other Hofstede indices are ‘power distance’, ‘masculinity’, ‘uncertainty avoidance’ and ‘indulgence’. As it turns out, neither of them directly affects secondary schooling, nor is there an interaction effect between ‘pronoun drop’ and any of these measures. ‘Pronoun drop’ does not affect schooling through any of these measures either (results not reported here).
of dimensions of contemporary culture: ‘embeddedness’ and ‘autonomy’ (Licht et al. 2007), ‘trust and respect’ (Tabellini 2008), ‘individualism’ (Klasing 2013, Gorodnichenko and Roland 2017) and ‘power distance’ (Klasing 2013). Unfortunately, apart from the downside that these papers use Kashima and Kashima’s (1998) problematic data (Davis and Abdurazokzoda 2016), they have several econometric issues. First, pronoun drop can be a valid instrument for at most one of these measures (Bazzi and Clemens 2013). Second, in several of these papers the first-stage $F$ statistic is below ten, indicating that pronoun drop is a weak instrument (Staiger and Stock 1997). Third, except for Gorodnichenko and Roland (2017), none of the papers provides statistical evidence that pronoun drop is exogenous.

In line with most of these papers, we also find ‘pronoun drop’ to be a weak instrument, for all four contemporary culture variables used here. This is evident not just from the first-stage $F$ statistic but also from Shea’s (1997) partial $R^2$ and Kleibergen and Paap’s (2006) rk LM statistic. As an example, Table A6 (in the online Supporting Information file) reports the results from IV regressions in which we use ‘pronoun drop’ as an instrument for ‘individualism’, the measure which is most likely to be relevant here. Furthermore, as in Table 5, ‘individualism’ is statistically insignificant in the IV regressions too – as are the other three contemporary variables in their IV regressions (the latter results are not reported here).

What is the upshot of all of these exercises involving contemporary culture variables? Most importantly, we find no evidence that any of them affect secondary schooling. Nor do we find evidence that ‘pronoun drop’ affects schooling indirectly, through contemporary culture. Rather, the results from these exercises reinforce our previous result that ‘pronoun drop’ appears to have a direct effect. The effect may be slightly smaller in countries with more long-term orientation. More intriguingly, the fact that the effect of ‘pronoun drop’ is smaller and less statistically significant when adding ‘individualism’ lends tentative support to our conjecture that pronoun drop languages may be an indicator of, and may perpetuate, ancient collectivism. Specifically, pronoun drop languages may pass on to their speakers the ancestral cultural norm that the individual has to subordinate to the collective and, more specifically, that education should be limited in order to avoid that young individuals reduce their commitment to the extended family and the state. As a result of such beliefs, which appear to be maintained by a language structure that accentuates the social embeddedness of the individual, educational investments in young people may remain limited. They may remain especially limited in girls because, as explained in section II, females’ contributions to maintaining the extended family are particularly important in collectivist societies.

How can an indicator of ancient collectivism affect contemporary school enrollment rates while related indicators of contemporary culture do not? Apart
from the fact that both types of indicators have measurement issues, there may be a fundamental reason for the answer to be in the affirmative. As explained previously, we think that rules allowing pronoun drop reflect and perpetuate deep-routed cultural values of collectivism that put comparatively little emphasis on the educational advancement of the young, especially girls. By contrast, contemporary values have become much more individualistic in recent decades, including in many traditionally collectivist societies. These more individualistic values foster and call for the education of the young. In all societies, traditional and modern values co-exist alongside each other. For example, in a seminal paper that uses World Values Survey data from the 1980s and 1990s on 65 societies and 75% of the world population, Inglehart and Baker (2000) provide evidence of both massive cultural change and the persistence of traditional values. They find economic development to be associated with shifts towards values that are increasingly rational, tolerant, trusting and participatory – values that support and call for education. But they also find that the cultural heritage of a society leaves an imprint on values that endures despite modernization. Thus, Inglehart and Baker’s (2000) findings support our interpretation of our results.

Finally, a brief comment on the estimates for the controls (Tables 2–4). Many of them accord with the previous literature. For example, we find longer life expectancy to have a positive effect and larger shares of both children and elderly to have negative effects on schooling. Additionally, girls’ education is negatively affected by Islam and positively by both urbanization and openness. Also in line with the previous literature, we find that economic freedom has a positive effect on boys as well as girls and that the legacies of both Spanish and French colonial education still adversely affected schooling in the respective former colonies in recent decades. Furthermore, we find evidence that schooling is countercyclical and that tropical climates affect schooling negatively. Both of these results are in line with the previous literature too. In contrast to Chen (2013) and Galor et al. (2017) though, we find no effect of future-time rules.

Previously, we have indicated that pronoun drop languages may affect schooling through public spending on education. However, in some unreported regressions we find no evidence for such an indirect effect, nor for a corresponding interaction effect. Therefore, we stick to the commonly used specification of including ‘public spending on education’ as a regular control variable. The fact that

9Pronoun drop rules are not perfect measures of ancient collectivism. For example, the extent to which ancient norms and values of collectivism were encoded in such rules and thus the extent to which they reflect ancient collectivism is likely to vary across languages. Furthermore, the strength of collectivist norms is also likely to have varied across groups of speakers of different languages in the distant past. Survey-based indicators such as the ones used for contemporary culture have a host of measurement issues of their own (e.g., Bertrand and Mullainathan 2001).
it turns out to be insignificant in this specification, too, is little surprising. Several other papers do not find a statistically significant effect of public education spending on secondary school enrollment either (e.g., Flug et al. 1998, Papagapitos and Riley 2009).

In some more supplementary regressions, we interact ‘pronoun drop’ with ‘political rights & civil liberties’ and, alternatively, with ‘economic freedom’. The intention here is to check whether the negative effect of pronoun drop is smaller in countries with more political or economic freedom. However, we find the respective interaction terms to be statistically insignificant (results not reported here).

Lastly, it should be noted that the large number of control variables we use ensures that ‘pronoun drop’ does not proxy for factors such as GDP per capita, public education spending, political or economic freedom, religion, or demographic or geographic conditions. As we have seen, it does not proxy for contemporary culture either.

V. CONCLUSION

According to our regression results, languages that allow to omit pronominal subjects have a negative effect on human capital. Specifically, speakers of such languages have a lower probability of having completed secondary or tertiary education, compared with speakers of languages that do not allow such an omission. Furthermore, countries where the dominant languages permit pronoun drop have lower secondary school enrollment rates. In both cases, the magnitude of the effect is substantial and slightly larger among females. Furthermore, we find contemporary culture not to attenuate the effect of pronoun drop languages, except very slightly so in the cases of contemporary individualism/collectivism and long-term orientation. Finally, we find pronoun drop languages to affect schooling directly rather than through contemporary culture. In our view, these languages reflect and perpetuate cultural norms of collectivism that have been formed in the distant past. While in many traditionally collectivist societies collectivist norms may be in retreat in contemporary culture, in such societies these ancient norms appear to live on and still adversely affect human capital today.

Although our regressions control for numerous factors and the results are robust, more research is clearly warranted. First and most importantly, the channels through which pronoun drop languages affect human capital need to be studied in detail. Second, the effect of these languages on other forms of human capital, from technical knowledge and skills to health, should to be analyzed too. As human capital is a key driver of economic and social development, a better understanding of the role played by such languages is an important issue for future research.
REFERENCES

Acemoglu, Daron (2005). Constitutions, Politics, and Economics: A Review Essay on Persson and Tabellini’s The Economic Effects of Constitutions, Journal of Economic Literature. 43(4): 1025-1048.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat and Romain Wacziarg (2003). Fractionalization, Journal of Economic Growth. 8(2): 155-194.

Baronchelli, Andrea, Vittorio Loreto and Andrea Puglisi (2015). Individual Biases, Cultural Evolution, and the Statistical Nature of Language Universals: The Case of Colour Naming Systems, arXiv. 1310.7782v2.

Baskaran, Thushyanthan and Zohal Hessami (2012). Public Education Spending in a Globalized World: Is there a Shift in Priorities across Educational Stages? International Tax and Public Finance. 19(5): 677-707.

Bazzi, Samuel and Michael A. Clemens (2013). Blunt Instruments: Avoiding Common Pitfalls in Identifying the Causes of Economic Growth, American Economic Journal: Macroeconomics. 5(2): 152-186.

Becker, Gary S. and H. Gregg Lewis (1973). On the Interaction between the Quantity and Quality of Children, Journal of Political Economy. 81(2, 2): S279-S288.

Bertinelli, Luisito and Benteng Zou (2008). Does Urbanization Foster Human Capital Accumulation? Journal of Developing Areas. 41(2): 171-184.

Bertrand, Marianne and Sendhil Mullainathan (2001). Do People Mean What They Say? Implications for Subjective Survey Data, American Economic Review, Papers and Proceedings. 91(2): 67-72.

Cavalli-Sforza, Luigi Luca (2000). Genes, Peoples, and Languages. New York: North Point.

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.

Christiansen, Morten H. and Simon Kirby (2003). Language Evolution: Consensus and Controversies, Trends in Cognitive Sciences. 7(7): 300-307.

Cooray, Arusha and Niklas Potrafke (2011). Gender Inequality in Education: Political Institutions or Culture and Religion? European Journal of Political Economy. 27(2): 268-280.

Darwin, Charles (1859). On the Origin of Species by Means of Natural Selection. London: Murray.

Davis, Lewis S. and Farangis Abdurazokzoda (2016). Language, Culture and Institutions: Evidence from a New Linguistic Dataset, Journal of Comparative Economics. 44(3): 541-561.

De Gregorio, José (1996). Borrowing Constraints, Human Capital Accumulation, and Growth, Journal of Monetary Economics. 37(1): 49-71.

Dryer, Matthew S. (2013). Expression of Pronominal Subjects, in: Matthew S. Dryer and Martin Haspelmath (eds.), World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at http://wals.info/chapter/101, accessed on 21 October 2016.

Evans, Nicholas (2013). Language Diversity as a Resource for Understanding Cultural Evolution, in: Peter J. Richerson and Morten H. Christiansen (eds.), Cultural Evolution: Society, Technology, Language, and Religion. Cambridge (MA): MIT Press: 233-268.

Evans, Nicholas and Stephen C. Levinson (2009). The Myth of Languages Universals: Language Diversity and Its Importance for Cognitive Science, Behavioral and Brain Sciences. 32(5): 429-492.

Everett, Caleb (2013). Linguistic Relativity: Evidence across Languages and Cognitive Domains. Berlin: De Gruyter Mouton.

Feldmann, Horst (2016a). Which Religions Still Affect Schooling? A Study of 143 Countries, Comparative Sociology. 15(4): 439-484.

Feldmann, Horst (2016b). The Long Shadows of Spanish and French Colonial Education, Kyklos. 69(1): 32-64.
Feldmann, Horst (2017). Economic Freedom and Human Capital Investment, *Journal of Institutional Economics*. 13(2): 421-445.

Flug, Karnit, Antonio Spilimbergo and Erik Wachtenheim (1998). Investment in Education: Do Economic Volatility and Credit Constraints Matter? *Journal of Development Economics*. 55(2): 465-481.

Forston, Jane G. (2011). Mortality Risk and Human Capital Investment: The Impact of HIV/AIDS in Sub-Saharan Africa, *Review of Economics and Statistics*. 93(1): 1-15.

Gallup, John Luke, Jeffrey D. Sachs and Andrew D. Mellinger (1999). Geography and Economic Development, *International Regional Science Review*. 22(2): 179-232.

Galar, Oded, Ömer Özak and Assaf Sarid (2017). Geographical Origins and Economic Consequences of Language Structures, Working Paper. Providence (RI): Brown University.

Gay, Victor, Daniel L. Hicks, Estefania Santacreu-Vasut and Amir Shoham (2018). Decomposing Culture: An Analysis of Gender, Language, and Labor Supply in the Household, *Review of Economics of the Household*. 16(4): 879-909.

Gorodnichenko, Yuriy and Gerard Roland (2017). Culture, Institutions, and the Wealth of Nations, *Review of Economics and Statistics*. 99(3): 402-416.

Hausman, Jerry A. (1978). Specification Tests in Econometrics, *Econometrica*. 46(6): 1251-1271.

Hayek, Friedrich A. (2010 [1942-44]). Scientism and the Study of Society, in: Friedrich A. Hayek, *Studies on the Abuse and Decline of Reason*, Collected Works, Vol. XIII. Chicago (IL): University of Chicago Press: 75-166.

Hayek, Friedrich A. (2014 [1965]). Kinds of Rationalism, in: Friedrich A. Hayek, *The Market and Other Orders*, Collected Works, Vol. XV. Chicago (IL): University of Chicago Press: 39-53.

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

Hofstede, Geert (2001). *Culture’s Consequences: Comparing Values, Behaviors, Institutions, and Organizations across Nations*, 2nd ed. Thousand Oaks (CA): Sage.

Hofstede, Geert (2014). Dimension Data Matrix. Available at http://www.geerthofstede.com, accessed 11 April 2014.

Inglehart, Ronald and Wayne E. Baker (2000). Modernization, Cultural Change, and the Persistence of Traditional Values, *American Sociological Review*. 65(1): 19-51.

Jäntti, Markus and Stephen P. Jenkins (2014). Income Mobility, in: Anthony B. Atkinson and François Bourguignon (eds.), *Handbook of Income Distribution*, Vol. 2A. Amsterdam: Elsevier: 807-935.

Johansson, Sverker (2005). *Origins of Language: Constraints on Hypotheses*. Amsterdam: Benjamins.

Kashima, Emiko S. and Yoshihisa Kashima (1998). Culture and Language: The Case of Cultural Dimensions and Personal Pronoun Use, *Journal of Cross-Cultural Psychology*. 29(3): 461-486.

Kashima, Yoshihisa and Emiko S. Kashima (2003). Individualism, GNP, Climate, and Pronoun Drop: Is Individualism Determined by Affluence and Climate, or Does Language Use Play a Role? *Journal of Cross-Cultural Psychology*. 34(1): 125-134.

Kirby, Simon, Mike Dowman and Thomas L. Griffiths (2007). Innateness and Culture in the Evolution of Language, *Proceedings of the National Academy of Sciences of the United States of America*. 104(12): 5241-5245.

Klasing, Mariko J. (2013). Cultural Dimensions, Collective Values and their Importance for Institutions, *Journal of Comparative Economics*. 41(2): 447-467.

Kleibergen, Frank and Richard Paap (2006). Generalized Reduced Rank Tests Using the Singular Value Decomposition, *Journal of Econometrics*. 133(1): 97-126.

Lake, David A. and Matthew A. Baum (2001). The Invisible Hand of Democracy: Political Control and the Provision of Public Services, *Comparative Political Studies*. 34(6): 587-621.

Levinson, Stephen C. (2003). Language and Mind: Let’s Get the Issues Straight! in: Dedre Gentner and Susan Goldin-Meadow (eds.), *Language in Mind: Advances in the Study of Language and Thought*. Cambridge (MA): MIT Press: 25-46.
Licht, Amir N., Chanan Goldschmidt and Shalom H. Schwartz (2007). Culture Rules: The Foundations of the Rule of Law and Other Norms of Governance, *Journal of Comparative Economics*. 35(4): 659-688.

Liu, Zhiqiang (2003). The Economic Impact and Determinants of Investment in Human and Political Capital in China, *Economic Development and Cultural Change*. 51(4): 823-849.

Long, J. Scott (1997). Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks (CA): Sage.

Lucy, John A. (1997). Linguistic Relativity, *Annual Review of Anthropology*. 26: 291-312.

Mavitsakalyan, Astghik (2015). Gender in Language and Gender in Employment, *Oxford Development Studies*. 43(4): 403-424.

Mavitsakalyan, Astghik and Clas Weber (2018). Linguistic Structures and Economic Outcomes, *Journal of Economic Surveys*. 32(3): 916-939.

McFadden, Daniel (1973). Conditional Logit Analysis of Qualitative Choice Behavior, in: Paul Zarembka (ed.), *Frontiers in Econometrics*. New York: Academic Press: 105-142.

McKelvey, Richard D. and William Zavoina (1975). A Statistical Model for the Analysis of Ordinal Dependent Variables, *Journal of Mathematical Sociology*. 4(1): 103-120.

Méndez, Fabio and Facundo Sepúlveda (2012). The Cyclicality of Skill Acquisition: Evidence from Panel Data, *American Economic Journal: Macroeconomics*. 4(3): 128-152.

Mincer, Jacob (1996). Economic Development, Growth of Human Capital, and the Dynamics of the Wage Structure, *Journal of Economic Growth*. 1(1): 29-48.

Norton, Seth W. and Annette Tomal (2009). Religion and Female Educational Attainment, *Journal of Money, Credit and Banking*. 41(5): 961-986.

Papagapitos, Agapitos and Robert Riley (2009). Social Trust and Human Capital Formation, *Economics Letters*. 102(3): 158-160.

Poterba, James M. (1997). Demographic Structure and the Political Economy of Public Education, *Journal of Policy Analysis and Management*. 16(1): 48-66.

Pütz, Martin and Marjolijn H. Verspoor (eds., 2000). *Explorations in Linguistic Relativity*. Amsterdam: Benjamins.

Santacreu-Vasut, Estefania, Amir Shoham and Victor Gay (2013). Do Female/Male Distinctions in Language Matter? Evidence from Gender Political Quotas, *Applied Economics Letters*. 20(5): 495-498.

Santacreu-Vasut, Estefania, Oded Shenkar and Amir Shoham (2014). Linguistic Gender Marking and Its International Business Ramifications, *Journal of International Business Studies*. 45(9): 1170-1178.

Sapir, Edward (1963). *Selected Writings of Edward Sapir in Language, Culture and Personality*, edited by David G. Mandelbaum. Berkeley (CA): University of California Press.

Schwartz, Shalom H. (2004). Mapping and Interpreting Cultural Differences around the World, in: Henk Vinken, Joseph Soeters and Peter Ester (eds.), *Comparing Cultures: Dimensions of Culture in a Comparative Perspective*. Leiden: Brill: 43-73.

Schwartz, Shalom H. (2006). A Theory of Cultural Value Orientations: Explication and Applications, *Comparative Sociology*. 5(2-3): 137-182.

Schwartz, Shalom H. (2008). *Cultural Value Orientations: Nature & Implications of National Differences*, Manuscript. Jerusalem: Hebrew University of Jerusalem.

Shea, John (1997). Instrument Relevance in Multivariate Linear Models: A Simple Measure, *Review of Economics and Statistics*. 49(2): 348-352.

Soares, Rodrigo R. (2005). Mortality Reductions, Educational Attainment, and Fertility Choice, *American Economic Review*. 95(3): 580-601.

Staiger, Douglas and James H. Stock (1997). Instrumental Variables Regression with Weak Instruments, *Econometrica*. 65(3): 557-586.

Tabellini, Guido (2008). Institutions and Culture, *Journal of the European Economic Association*. 6(2-3): 255-294.
HORST FELDMANN

Taubman, Paul (1989). Role of Parental Income in Educational Attainment, *American Economic Review*. 79(2): 57-61.

Thompson, Bill, Simon Kirby and Kenny Smith (2016). Culture Shapes the Evolution of Cognition, *Proceedings of the National Academy of Sciences of the United States of America*. 113(16): 4530-4535.

Triandis, Harry C. (2015). Collectivism and Individualism: Cultural and Psychological Concerns, in: James D. Wright (ed.), *International Encyclopedia of the Social & Behavioral Sciences*, 2nd ed., Vol. 4. Amsterdam: Elsevier: 206-210.

Trostel, Philip A. (2002). Should Education Be Publicly Provided? *Bulletin of Economic Research*. 54(4): 373-391.

Veall, Michael R. and Klaus F. Zimmermann (1996). Pseudo-R2 Measures for Some Common Limited Dependent Variable Models, *Journal of Economic Surveys*. 10(3): 241-259.

Whorf, Benjamin Lee (2012 [1956]). *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, 2nd ed., edited by John B. Carroll, Stephen C. Levinson and Penny Lee. Cambridge (MA): MIT Press.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table A1. Individual-level countries
Table A2. Individual-level variables
Table A3. Country-level variables
Table A4. Country-level countries
Table A5. Country-level correlation matrix: pronoun drop and contemporary culture
Table A6. Country-level IV regressions

SUMMARY

This paper empirically studies the human capital effects of grammatical rules that permit speakers to drop a personal pronoun when used as a subject of a sentence. By de-emphasizing the significance of the individual, such languages may perpetuate ancient values and norms that give primacy to the collective, inducing governments and families to invest relatively little in education because education usually increases the individual’s independence from both the state and the family and may thus reduce the individual’s commitment to these institutions. Carrying out both an individual-level and a country-level analysis, the paper indeed finds negative effects of pronoun-drop languages. The individual-level analysis uses data on 114,894 individuals from 75 countries over 1999-2014. It establishes that speakers of such languages have a lower probability of having completed secondary or tertiary education, compared with speakers of languages that do not allow pronoun drop. The country-level analysis uses data from 101 countries over 1972-2012. Consistent with the individual-level analysis, it finds that countries where the dominant languages permit pronoun drop have lower secondary school enrollment rates. In both cases, the magnitude of the effect is substantial, particularly among females.