The Effect of Language on Economic Behavior: Evidence from Savings Rates, Health Behaviors, and Retirement Assets

Online Appendix

M. Keith Chen
December, 2012

1 Alternate Measures of FTR Structure

The analyses in this appendix investigate the sensitivity of my results to alternate ways of measuring a language's FTR structure. In section 4.2 and in the published appendix of the main paper I discuss two alternatives to the strong vs. weak FTR dichotomy. Investigating how my results change when these alternative measures are used can be thought of as a robustness test of the binary dichotomy I use in the main paper.

1.1 Regressions with Online Language Measures

Section 4.2 and the main appendix of the paper describe a measure of FTR strength based on word-frequency analysis of text retrieved from online full-sentence weather forecasts. As of the writing of the main paper, this analysis covers 39 languages which are well-represented on the internet.

Table 1 in the main appendix reports two measures of how frequently a weather reports grammatically marks future time. "Verb ratio" counts the number of verbs which are grammatically future-marked, divided by the total number of future-referring verbs. In other words: in online weather forecasts in a language, what share of verbs about future weather are marked as future-referring? Similarly, "sentence ratio" asks: what share of sentences regarding future weather contain a grammatical future marker? In some languages (Arabic for example), often a sentence with multiple verbs will grammatically mark only the first as future-regarding. Differences between languages in rules like these lead to variation between verb and sentence ratios.

1.1.1 Regressions with Online Language Measures

While the set of languages codable in this way is limited to those which are well represented on the searchable internet, it is extensive enough that both the OECD and SHARE results I report can be run using either ratio instead of the binary weak vs. strong FTR measure. Online Appendix Tables 1, 2, and 3 report the results of these regressions.

Table 1 reports regressions of OECD savings rates on our two online language measures and numerous economic and demographic controls commonly found in studies of national savings. These regressions are identical in form to those reported in Table 10 of the main paper. Please see the main paper for details on both the estimating equation and details on the controls included in these regressions.

Online Appendix Table 1: GDSRs in the OECD and Online Language Measures

	(1) $GDSR_t$	(2) $GDSR_t$	(3) $GDSR_t$	(4) $GDSR_t$	(5) $GDSR_t$	(6) $GDSR_t$
Sentence Ratio	-5.446		-6.531		-6.124	
	[1.789]**		[2.029]**		[1.579]**	
Verb Ratio		-6.131		-6.987		-6.774
		[1.911]**		[2.139]**		[1.610]**
$PCGDP_{t-1}$ / $PCGDP_t$	-32.864	-32.528	-43.532	-42.909	-32.454	-32.441
	[8.140]**	[7.971]**	[14.583]**	[14.221]**	[12.025]*	[11.875]*
CAGR	-0.118	-0.127	0.032	0.001	0.010	-0.011
	[0.104]	[0.102]	[0.209]	[0.205]	[0.173]	[0.170]
$Unemployment_t$ (%)	-0.462	-0.44	-0.207	-0.209	-0.301	-0.296
	[0.167]**	[0.163]*	[0.153]	[0.149]	[0.179]	[0.178]
Old_t (%)	-1.162	-1.117	-1.235	-1.154	-1.327	-1.229
	[0.339]**	[0.328]**	[0.366]**	[0.351]**	[0.370]**	[0.361]**
$Young_t$ (%)	-0.544	-0.508	-0.364	-0.339	-0.203	-0.163
	[0.190]**	[0.187]*	[0.275]	[0.266]	[0.215]	[0.213]
1 / $PCGDP_t$			-87.681	-78.234	-115.33	-110.81
			[59.121]	[58.455]	[45.840]*	[45.384]*
Soc Sec_t (%GDP / Old)			-3.215	-3.178	-4.638	-4.476
			[2.285]	[2.349]	[2.678]	[2.654]
Protestant					-3.808	-3.941
					[1.372]*	[1.361]**
Dist from Equator					2.867	2.660
					[1.520]	[1.491]
Corresponding Coef.	-5.272	-5.272	-5.245	-5.245	-5.730	-5.730
on Strong FTR	[1.798]**	[1.798]**	[1.948]*	[1.948]*	[1.454]**	[1.454]**
Observations	841	841	564	564	564	564
R-squared	0.43	0.45	0.49	0.49	0.58	0.59

Regressions are OLS regressions where the dependent variable is a country's Gross Domestic Savings Rate in year t. Observations are for OECD countries from 1970 to 2009. Protestant is a binary variable that measures if the country is majority protestant or not. Robust standard errors are reported in brackets and clustered at the country level. * significant at 5%; ** significant at 1%.

Similar to the regressions from Table 10 in the main paper, these regressions suggests that countries whose languages never grammaticalize future-time reference save on average about six percentage points more than those which mark FTR 100% of the time.

For the sake of comparison, Table 1 also lists the coefficient on Strong FTR for each regression when my original measure of FTR used. The results I obtain when substituting in either the sentence or verb ratio are nearly identical (both quantitatively and statistically) to the corresponding coefficients on Strong FTR. This suggests the results I report in the main paper are robust to different ways of measuring languages' FTR structure.

Table 2 reports regressions of accumulated retirement assets in the SHARE on our two online language measures. These regressions are identical in form to those reported in Table 6 of the main paper. Please see the main paper for details on both the estimating equation and details on the controls included in these regressions.

Online Appendix Table 2: Ret. Assets in the SHARE and Online Measures

	(1) $IHS\left(\frac{RA}{DI}\right)$	(2) $IHS\left(\frac{RA}{DI}\right)$	(3) $IHS\left(\frac{RA}{DI}\right)$	(4) $IHS\left(\frac{RA}{DI}\right)$	(5) $IHS\left(\frac{RA}{DI}\right)$	(6) $IHS\left(\frac{RA}{DI}\right)$
Sentence Ratio	-0.400		-0.396		-0.366	
	[0.017]**		[0.047]**		[0.077]**	
Verb Ratio		-0.408		-0.404		-0.373
		[0.017]**		[0.048]**		[0.078]**
Fixed Effects:						
Age	Yes	Yes	Yes	Yes	Yes	Yes
Country × Wave	Yes	Yes	Yes	Yes	Wave	Wave
Income	No	No	Yes	Yes	Yes	Yes
Education	No	No	Yes	Yes	Yes	Yes
Married × Num Chil	No	No	Yes	Yes	Yes	Yes
All FEs Interacted	Yes	Yes	Yes	Yes	Yes	Yes
Countries	All	All	All	All	BE & CH	BE & CH
Corresponding Coef.	-0.390	-0.390	-0.386	-0.386	-0.356	-0.356
on Strong FTR	[0.017]**	[0.017]**	[0.047]**	[0.047]**	[0.079]**	[0.079]**
Observations	39,665	39,665	39,350	39,350	5,937	5,937
F stat	547.74	551.77	70.40	70.76	22.84	22.86

Regressions are fixed-effect OLS regressions where the dependent variable is the inverse-hyperbolic sine of net household retirement assets divided by average national disposable income. Immigrant households are excluded from all regressions. Robust standard errors are reported in brackets; all regressions are clustered at the country level except regression 5, which is clustered at the household level. * significant at 5%; ** significant at 1%.

Regressions 1 through 6 show my predicted effect carries through to using online language FTR measures; moving from a language which does not grammaticalize future-time reference to one that marks it 100% of the time leads households accumulating around 39% less by the time they retire. These regressions are largely identified by the fact that Belgium has large Flemish (weak-FTR) and French (strong-FTR) speaking populations, and Switzerland has large German (weak-FTR), and French, Italian, and Romansh (strong-FTR) speaking populations.

For the sake of comparison, Table 2 also lists the coefficient on strong-FTR for each regression when that is the measure of FTR used. The results I obtain when substituting in either the sentence or verb ratio are nearly identical (both quantitatively and statistically) to the corresponding coefficients on Strong FTR. This suggests the results I report in the main paper are relatively robust to the specification of strong and weak FTR.

Table 3 reports regressions of health behaviors in the SHARE on our two online language measures and a large number of demographic controls. These regressions are identical in form to those reported in Table 8 of the main paper. Please see the main paper for details on both the estimating equation and details on the controls included in these regressions.

Online Appendix Table 3: Health Behaviors in the SHARE and Online Measures

	(1)	(2)	(3)	(4)	(5)	(6)
	Smoked	Smoked	Phy Act	Phy Act	Obesity	Obesity
Sentence Ratio	1.248		0.704		1.135	
	[0.042]**		[0.026]**		[0.006]**	
Verb Ratio		1.254		0.699		1.138
		[0.043]**		[0.026]**		[0.006]**
Full set of FEs						
from reg 4 table 6	Yes	Yes	Yes	Yes	Yes	Yes
All FEs Interacted	Yes	Yes	Yes	Yes	Yes	Yes
Corresponding Coef.	1.241	1.241	0.709	0.709	1.131	1.131
on Strong FTR	[0.042]**	[0.042]**	[0.025]**	[0.025]**	[0.007]**	[0.007]**
Observations	15,750	15,750	9,135	9,135	11,958	11,958

Regressions are fixed-effect (or conditional) logistic regressions with coefficients reported as odds ratios. The dependent variables are: having smoked daily for a year or more, engaging in regular physical activity, and medically obesity. Immigrants are excluded from all regressions. Robust standard errors are reported in brackets; all regressions are clustered at the country level. * significant at 5%; ** significant at 1%.

Regressions 1 and 2 indicate that moving from a language which does not grammaticalize future-time reference to one that marks it 100% of the time leads to a 25% higher probability of having ever smoked (daily for a year or more). This is consistent with my main findings on savings if the decision to smoke trades off immediate benefits versus future health costs. Regressions 3, 4, 5, and 6 show similar effects for both self-reported physical activeness and measured obesity.

For the sake of comparison, Table 3 also lists the coefficient on strong-FTR for each regression when that is the measure of FTR used. The results I obtain when substituting in either the sentence or verb ratio are nearly identical (both quantitatively and statistically) to the corresponding coefficients on Strong FTR. This suggests the results I report in the main paper are relatively robust to the specification of strong and weak FTR.

2 Regressions with Alternative Typological Language Measures

Section 4.2 of the paper describes two alternative typological distinctions in addition to the strong vs. weak FTR classification I examine in the main paper. **Any FTR** is a weaker criterion which marks the presence of *any grammatical marking* of future events in a language, even if infrequently used. This would include both *inflectional* markers (like the future-indicating suffixes in Romance languages) or *periphrastic* markers (like the English auxiliary 'will'). Mandarin, Finnish, and Estonian are examples of languages that lack either type of future markers. **Inflectional FTR** is a stronger criterion which marks the presence of an inflectional future tense. These alternative criterion satisfy:

$$\text{Any Gr FTR} \supset \text{Weak FTR} \supset \text{Strong FTR} \overset{?}{\supset} \text{Inflectional FTR}, \tag{1}$$

with the first and second inclusions being logically necessary, and the third representing a typological regularity for which I do not have a counterexample.

A natural hypothesis would be that as we move from weaker to stronger measures of a language's FTR structure, the effects I measure in the main paper would strengthen. Unfortunately, this divides languages into 4 sets rather than the 2 defined by strong and weak FTR, which lowers the power of the regressions in the paper and leads to identification off of very narrow sets of languages. For example, once country fixed effects eliminate cross-country variation, Estonian is the only remaining European language with no grammaticalized FTR, and Mandarin is the only remaining Asian language.

However, it is possible to include all three criteria as nested effects in the broader cross-country savings regressions I run. Online Appendix Table 4 presents regressions with these nested effects added to cross-country savings regressions in the World Values Survey (Table 12 in the main paper). Please see the main paper for details on both the estimating equation and details on the controls included in these regressions.

Online Appendix Table 4: Savings Rates in the WVS and Nested FTR Measures

	(1) $GDSR_t$	(2) $GDSR_t$	(3) $GDSR_t$	(4) $GDSR_t$	(5) $GDSR_t$	(6) $GDSR_t$
Any FTR		-5.752		-3.716		-2.526
		[6.306]		[4.254]		[3.827]
Strong FTR	-15.545	-12.566	-12.253	-10.233	-11.328	-8.836
	[4.814]**	[4.802]*	[3.337]**	[3.939]*	[3.320]**	[3.871]*
Inflectional FTR		-1.032		-0.815		-2.828
		[6.520]		[4.959]		[4.162]
$PCGDP_{t-1}$ / $PCGDP_t$	19.905	22.469	15.108	16.564	15.616	11.208
	[28.120]	[29.389]	[25.197]	[26.142]	[23.366]	[22.869]
Old_t (%)	-1.718	-1.571	-1.916	-1.807	-2.112	-1.881
	[0.839]*	[1.043]	[0.730]*	[0.860]*	[0.687]**	[0.769]*
$Young_t$ (%)	-0.736	-0.632	-0.813	-0.737	-0.891	-0.728
	[0.498]	[0.710]	[0.501]	[0.620]	[0.512]	[0.587]
French Legal Origin	-7.676	-7.143	-3.302	-2.929	-7.578	-7.748
	[2.843]**	[3.658]	[2.828]	[3.448]	[4.887]	[4.778]
German Legal Origin	-9.937	-9.951	-6.735	-6.702	-11.716	-12.253
	[6.790]	[6.308]	[4.980]	[4.681]	[4.828]*	[4.980]*
Scandanavian Lgl Or	-7.430	-7.376	-3.196	-3.13	-6.432	-6.595
	[7.248]	[7.644]	[5.326]	[5.570]	[5.355]	[5.825]
1 / $PCGDP_t$	-4.455	-4.721	-4.819	-5.001	-5.102	-5.541
	[1.726]*	[2.248]*	[1.781]**	[2.067]*	[1.766]**	[2.003]**
$Unemployment_t$ (%)			-0.724	-0.727	-0.587	-0.573
			[0.193]**	[0.196]**	[0.225]*	[0.219]*
Real Interest $Rate_t$			-0.199	-0.193	-0.219	-0.220
			[0.108]	[0.107]	[0.092]*	[0.091]*
Legal Rights Index					-0.899	-1.144
					[0.999]	[0.892]
$Trust_t$					2.947	1.943
					[9.244]	[9.447]
Family is $Important_t$					47.163	48.469
					[15.877]**	[16.973]**
Observations	120	120	113	113	113	113
R-squared	0.32	0.33	0.51	0.51	0.55	0.56

Regressions are OLS regressions where the dependent variable is a country's Gross Domestic Savings Rates in year t. Observations are for the countries in the WVS countries over three waves, from 1994 to 2008. Robust standard errors are reported in brackets and clustered at the country level. * significant at 5%; ** significant at 1%.

These cross-country regressions suggest that as a language increasingly requires a grammatical separation of present and future events, countries which speak those languages tends to save less. While there is not enough variation to separate each level of additional grammatical FTR, results are broadly consistent with our findings when focussing the primary strong vs. weak FTR dimension.

Online Appendix Table 5 lists every language included in this study, and provides information about its family, genus, and whether it is strong of weak FTR.

Online Appendix Table 5: Coded Languages and FTR Values

Language	Family	Genus	FTR
Afrikaans	Indo-European	Germanic	Strong
Akan	Niger-Congo	Kwa	Strong
Alawa	Australian	Maran	Strong
Albanian	Indo-European	Albanian	Strong
Amharic	Afro-Asiatic	Semitic	Weak
Arabic	Afro-Asiatic	Semitic	Strong
Armenian	Indo-European	Armenian	Strong
Azari	Altaic	Turkic	Strong
Azerbaijani	Altaic	Turkic	Strong
Bandjalang	Australian	Pama-Nyungan	Strong
Bambara	Niger-Congo	Western Mande	Weak
Basque	Basque	Basque	Strong
Belorussian	Indo-European	Slavic	Strong
Bemba	Niger-Congo	Bantoid	Strong
Bengali	Indo-European	Indic	Strong
Beja	Afro-Asiatic	Beja	Weak
Bosnian	Indo-European	Slavic	Strong
Bulgarian	Indo-European	Slavic	Strong
Cantonese	Sino-Tibetan	Chinese	Weak
Catalan	Indo-European	Romance	Strong
Cebuano	Western Malayo-Polynesian	Meso-Philippine	Weak
Chaha	Afro-Asiatic	Semitic	Strong
Chichewa	Niger-Congo	Bantoid	Strong
Croatian	Indo-European	Slavic	Strong
Czech	Indo-European	Slavic	Strong
Dagbani	Niger-Congo	Gur	Strong
Danish	Indo-European	Germanic	Weak
Dutch	Indo-European	Germanic	Weak
Dyula	Niger-Congo	Western Mande	Weak
English	Indo-European	Germanic	Strong
Estonian	Finno-Ugric	Finnic	Weak
Ewe	Niger-Congo	Kwa	Strong
Finnish	Finno-Ugric	Finnic	Weak
Flemish	Indo-European	Germanic	Weak
French	Indo-European	Romance	Strong
Frisian	Indo-European	Germanic	Weak
Fula	Niger-Congo	Northern Atlantic	Strong
Gamo	Afro-Asiatic	North Omotic	Strong
Galician	Indo-European	Romance	Strong

Online Appendix Table 5: Coded Languages and FTR Values (Continued)

Language	Family	Genus	FTR
Georgian	Kartvelian	Kartvelian	Strong
German	Indo-European	Germanic	Weak
Greek	Indo-European	Greek	Strong
Guarani	Tupian	Tupi-Guarani	Strong
Gujarati	Indo-European	Indic	Strong
Hakka	Sino-Tibetan	Chinese	Weak
Hausa	Afro-Asiatic	West Chadic	Strong
Hawaiian	Eastern Malayo-Polynesian	Oceanic	Weak
Hebrew	Afro-Asiatic	Semitic	Strong
Hindi	Indo-European	Indic	Strong
Hungarian	Finno-Ugric	Ugric	Strong
Icelandic	Indo-European	Germanic	Weak
Igbo	Niger-Congo	Igboid	Strong
Irish	Indo-European	Celtic	Strong
Isekiri	Niger-Congo	Defoid	Strong
Indonesian	Western Malayo-Polynesian	Sundic	Weak
Italian	Indo-European	Romance	Strong
Japanese	Japanese	Japanese	Weak
Javanese	Western Malayo-Polynesian	Sundic	Weak
Kammu	Austro-Asiatic (Mon-Khmer)	Palaung-Khmuic	Strong
Kannada	Dravidian	Southern Dravidian	Strong
Karaim	Altaic	Turkic	Strong
Kongo	Niger-Congo	Bantoid	Weak
Korean	Korean	Korean	Strong
Kikuyu	Niger-Congo	Bantoid	Weak
Kurdish	Indo-European	Iranian	Strong
Latvian	Indo-European	Baltic	Strong
Lingala	Niger-Congo	Bantoid	Strong
Lithuanian	Indo-European	Baltic	Strong
Lozi	Niger-Congo	Bantoid	Strong
Luba	Niger-Congo	Bantoid	Strong
Luganda	Niger-Congo	Bantoid	Strong
Luxembourgish	Indo-European	Germanic	Weak
Malay	Western Malayo-Polynesian	Sundic	Weak
Maltese	Afro-Asiatic	Semitic	Strong
Macedonian	Indo-European	Slavic	Strong
Mandarin	Sino-Tibetan	Chinese	Weak
Maori	Western Malayo-Polynesian	Oceanic	Weak
Moldavian	Indo-European	Romance	Strong
Montenegrin	Indo-European	Slavic	Strong
Moore	Niger-Congo	Gur	Strong
Norwegian	Indo-European	Germanic	Weak
Oromo	Afro-Asiatic	Cushitic	Weak

Online Appendix Table 5: Coded Languages and FTR Values (Continued)

Language	Family	Genus	FTR
Panjabi	Indo-European	Indic	Strong
Persian	Indo-European	Iranian	Strong
Polish	Indo-European	Slavic	Strong
Portuguese (EU)	Indo-European	Romance	Strong
Portuguese (BR)	Indo-European	Romance	Weak
Quechua	Quechuan	Quechuan	Strong
Romanian	Indo-European	Romance	Strong
Romansh	Indo-European	Romance	Strong
Russian	Indo-European	Slavic	Strong
Serbian	Indo-European	Slavic	Strong
Slovak	Indo-European	Slavic	Strong
Slovene	Indo-European	Slavic	Strong
Soddo	Afro-Asiatic	Cushitic	Weak
Sotho (Northern)	Niger-Congo	Bantoid	Strong
Seraiki	Indo-European	Indic	Strong
Sesotho	Niger-Congo	Bantoid	Strong
Sidamo	Afro-Asiatic	Cushitic	Weak
Spanish	Indo-European	Romance	Strong
Sumatranese	Western Malayo-Polynesian	Sundic	Weak
Sundanese	Western Malayo-Polynesian	Sundic	Weak
Swati	Niger-Congo	Bantoid	Strong
Swedish	Indo-European	Germanic	Weak
Swahili	Niger-Congo	Bantoid	Strong
Swiss French	Indo-European	Romance	Strong
Swiss German	Indo-European	Germanic	Weak
Swiss Italian	Indo-European	Romance	Strong
Tagalog	Western Malayo-Polynesian	Meso-Philippine	Strong
Tamil	Dravidian	Southern Dravidian	Strong
Tenyer	Niger-Congo	Gur	Strong
Thai	Tai-Kadai	Kam-Tai	Strong
Tigrinya	Afro-Asiatic	Semitic	Strong
Tsonga	Niger-Congo	Bantoid	Strong
Tswana	Niger-Congo	Bantoid	Strong
Turkish	Altaic	Turkic	Strong
Ukrainian	Indo-European	Slavic	Strong
Urdu	Indo-European	Indic	Strong
Uzbek	Altaic	Turkic	Strong
Venda	Niger-Congo	Bantoid	Strong
Vietnamese	Austro-Asiatic (Mon-Khmer)	Viet-Muong	Strong
Wolaytta	Afro-Asiatic	North Omotic	Strong
Wolof	Niger-Congo	Northern Atlantic	Strong
Xhosa	Niger-Congo	Bantoid	Strong
Yoruba	Niger-Congo	Defoid	Weak
Zulu	Niger-Congo	Bantoid	Strong