Supplementary Materials for "Enculturation Trajectories: Language, Cultural Adaptation, and Individual Outcomes in Organizations"

1 Linguistic Inquiry and Word Count (LIWC)

Table SM1.1: Linguistic Inquiry and Word Count (LIWC)

Category	Examples	Words In Category
Total function words		464
Total pronouns	I, them, itself	116
Personal pronouns	I, them, her	70
1st pers singular	I, me, mine	12
1st pers plural	We, us, our	12
2nd person	You, your, thou	20
3rd pers singular	She, her, him	17
3rd pers plural	They, their, they'd	10
Impersonal pronouns	It, it's, those	46
Articles	A, an, the	3
Common verbs	Walk, went, see	383
Auxiliary verbs	Am, will, have	144
Past tense	Went, ran, had	145
Present tense	Is, does, hear	169
Future tense	Will, gonna	48
Adverbs	Very, really, quickly	69
Prepositions	To, with, above	60
Conjunctions	And, but, whereas	28
Negations	No, not, never	57
Quantifiers	Few, many, much	89
Numbers	Second, thousand	34
Swear words	Damn, piss, fuck	53
Social processes	Mate, talk, they, child	455
Family	Daughter, husband, aunt	64
Friends	Buddy, friend, neighbor	37
Humans	Adult, baby, boy	61
Affective processes	Happy, cried, abandon	915
Positive emotion	Love, nice, sweet	406
Negative emotion	Hurt, ugly, nasty	499
Anxiety	Worried, fearful, nervous	91
Anger	Hate, kill, annoyed	184
Sadness	Crying, grief, sad	101
Cognitive processes	cause, know, ought	730
Insight	think, know, consider	195

continued ...

Table SM1.1 (continued)

Category	Examples	Words In Category
Causation	because, effect, hence	108
Discrepancy	should, would, could	76
Tentative	maybe, perhaps, guess	155
Certainty	always, never	83
Inhibition	block, constrain, stop	111
Inclusive	And, with, include	18
Exclusive	But, without, exclude	17
Perceptual processes	Observing, heard, feeling	273
See	View, saw, seen	72
Hear	Listen, hearing	51
Feel	Feels, touch	75
Biological processes	Eat, blood, pain	567
Body	Cheek, hands, spit	180
Health	Clinic, flu, pill	236
Sexual	Horny, love, incest	96
Ingestion	Dish, eat, pizza	111
Relativity	Area, bend, exit, stop	638
Motion	Arrive, car, go	168
Space	Down, in, thin	220
Time	End, until, season	239
Work	Job, majors, xerox	327
Achievement	Earn, hero, win	186
Leisure	Cook, chat, movie	229
Home	Apartment, kitchen, family	93
Money	Audit, cash, owe	173
Religion	Altar, church, mosque	159
Death	Bury, coffin, kill	62
Assent	Agree, OK, yes	30
Nonfluencies	Er, hm, umm	8
Fillers	Blah, Imean, youknow	9

Accessed on May 8, 2015 from `http://www.liwc.net/descriptiontable1.php`

2 Data Processing

The electronic mail corpus includes 10,236,668 distinct messages exchanged over a period of five years, between 2009 and 2014. Before producing our measure of cultural fit, we applied a set of procedures to clean and structure the textual data contained in these messages. We began by removing all message headers and footers, eliminating concatenated messages from earlier iterations in the email thread, and extracting non-conversational metadata such as email addresses and timestamps. We mapped multiple email aliases of the same person

to a unique person identifier and discarded messages that were missing email addresses or timestamps (these constituted less than 0.001% of messages). We also discarded all messages exchanged with outside parties and with the firm's lawyers. We then segmented the data into monthly windows, tokenized the text into unique word stems, and mapped these tokens to LIWC categories (Pennebaker et al. 2007). We enumerated the frequency over LIWC categories for each person's incoming and outgoing monthly messages, respectively, and then normalized by person. To reduce the measure's susceptibility to outliers, we included only individual-month observations for individuals who sent and received a minimum of 20 emails per month. We then applied the Jensen-Shannon divergence-based method for calculating cultural fit to these structured data. Overall, this procedure resulted in 10,924 person-month observations.

Table SM2.1: Descriptive Statistics

	Observations	Mean	Standard Deviation	Min.	Max.
Individual Attributes					
Age (at time of entry)	601	33.2	9.71	19.8	66.8
Tenure (months)	601	19.6	15.5	1	89
Manager	601	0.240	0.427	0	1
Female	601	0.333	0.471	0	1
Person-Month Observations					
Cultural Fit	9885	2.101	0.452	0.228	3.39
No. emails received[†]	9885	1411.68	1255.60	20	21702
No. emails sent	9885	374.48	343.63	20	2610
Events					
Promotion to manager	118				
Departures (total)	224				
Voluntary Exit	89				
Involuntary Exit	135				

[†]On average, an email has more than one recipient. We count an email with multiple recipients as sent only once, hence the difference between total number of emails sent and received.

After producing the cultural fit measures, for the purpose of the multivariate analyses described below, we further excluded observations for the 36 interns included in the data (for whom departure date was predetermined) and for individuals whose demographic data (gender, age, or date of arrival/departure) was missing. This process excluded 1,039 person-month observations (which constituted 9.51% of the sample). Overall, our data processing procedure resulted in 9,885 person-month observations, comprising 601 full-time employees, which formed the basis of the analyses reported in the main text and detailed below. Descriptive statistics are provided in Table SM2.1.

3 Multivariate Analyses

We conducted a variety of multivariate analyses to model the relationship between cultural fit and individual outcomes, as reported in the main text. In this appendix we describe the models in full and report estimated coefficients.

Hazard of Promotion

Figure 2 plots the cumulative hazards of being promoted to a managerial position, or exiting involuntarily, estimated by two separate Cox proportional hazard models with time-varying predictors (Cox 1972). The hazard function has the form:

$$P(t|X) = \lambda_0(t) \exp(X\beta) \tag{SM3.1}$$

where, $P(t|X)$ is the hazard (probability) at time t that an individual will experience an event (promotion) that excludes her from the risk set, X is a set of explanatory variables and β is a set of estimated coefficients. The exponentiated coefficients can be interpreted as hazard ratios, or the ratio between the hazard rates of an event when the explanatory variable increases by 1 unit.

In Figure 2 we report the cumulative hazards of promotion to managerial position (Model 1), and of voluntary exit (Model 2), as estimated by the model in eq. SM3.1. The estimated coefficients are reported as hazard ratios in Table SM3.1. An increase in 1 unit of cultural fit (corresponding to 1 standard deviation, given that the measure is standardized) increases the hazard of promotion by more than 44%, and reduces the hazard of involuntary exit by 65%, suggesting that our measure of cultural fit is strongly predictive of individual attainment in the organization. We included individual (age, gender and managerial status in model 2) and organizational (department indicators) variables as controls. The hazard functions plotted in Figure 2 are calculated for mean values of the control variables and for the median, 5^{th} and 95^{th} percentiles of cultural fit.

Cultural Fit by Time Elapsed

Figure 3 reports a simple ordinary least squares model of the form

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_0 x^3 + \epsilon \tag{SM3.2}$$

where Y is cultural fit and x is time elapsed (in months) since joining the firm. Table SM3.2 reports the coefficients estimated by this model. The inset in Figure 3 plots the 3-month smoothed cultural fit values for 30 randomly selected individuals who are observed in the dataset for less than 36 months.

Cultural Fit by Tenure

Next, we estimated the different enculturation trajectories for those who remained in the organization and for those who departed either voluntarily or involuntarily. We observe departed employees throughout their tenures at the firm and can therefore model their trajectories in straightforward manner. Departed employees vary in the length of their tenure at the firm. We assume that different people follow different rates of enculturation and, as we explain in the main text, consequently standardized time by employees' tenure at the firm, notated as τ_i.

As illustrated in Figure 3 (inset), there is great heterogeneity in individuals' initial and peak levels of cultural fit. To account for this variability, we estimated a simple individual fixed-effects model of the form:

Table SM3.1: Cox Proportional Hazard Model of Attainment

	(1) Promotion	(2) Involuntary Exit
Cultural Fit	1.444**	0.445***
	(3.03)	(-6.85)
Age	1.324*	1.260*
	(2.35)	(2.53)
Age^2	0.997*	0.998
	(-2.15)	(-1.79)
Female	0.950	1.196
	(-0.21)	(0.71)
Manager		1.087
		(0.19)
Department Controls	Yes	Yes
N	8270	9885
χ^2	110.240	95.936
Log-Likelihood	-523.068	-358.065

Exponentiated coefficients; t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table SM3.2: OLS of Cultural Fit During First 36 Months

	(1) Cultural Fit
Months	0.056***
	(5.80)
$Months^2$	-0.002***
	(-3.70)
$Months^3$	0.00004**
	(3.11)
Intercept	-0.365***
	(-9.97)
N	9044
R^2	0.024

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

$$Y_{it} = X_{it}\beta + \alpha_i + \epsilon_{it} \tag{SM3.3}$$

where Y_{it} is cultural fit for individual i observed at time t, X_{it} are individual-time observed predictors, β are estimated coefficients, α_i is the unobserved time-invariant individual fixed effect, and ϵ_{it} is the error term. This model assumes that all individuals follow the same trajectory but vary in initial levels of cultural fit (i.e., their cultural adaptation functions have similar shapes but different intercepts). We included τ_i and τ_i^2 among the predictors in X to account for the curvilinearity of cultural adaptation trajectories (as implied by Figure 3).

We estimated two separate models, one for voluntarily departed and one for involuntarily departed employees. (Because departure type is time-invariant, we could not estimate a single individual fixed-effects model with interaction terms for departure type). We included time-varying controls (namely, department and managerial status). The estimated coefficients are reported in Table SM3.3. Panel C of Figure 4 plots the marginal effects of tenure on cultural fit, as well as their 95% confidence intervals, as estimated by these models.

Table SM3.3: Individual Fixed-Effects OLS Models of Cultural Fit

	(1)	(2)
	Involuntary Exit	Voluntary Exit
Tenure	0.696	1.302***
	(1.85)	(3.99)
Tenure^2	-0.829*	-1.313***
	(-2.50)	(-4.39)
Manager	0.060	0.049
	(0.75)	(0.59)
Intercept	-0.347***	-0.388***
	(-3.65)	(-4.76)
Department Controls	Yes	Yes
N	1119	961
R^2	0.655	0.710

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Individual fixed-effects models account for unobserved heterogeneity across individuals and therefore mitigate concerns about omitted variable bias (Greene 2012). In other words, these models estimate cultural fit trajectories net of fixed individual traits, thus addressing endogeneity related to time-invariant individual attributes (namely, differences in cultural fit trajectories that are explained by stable differences among individuals, rather than by tenure). Because these models are estimated separately for different exit types, they do not reliably estimate differences between exit types. Moreover, it is unclear whether these

trajectories differ from those of individuals who have not left the organization. Unlike with departed employees, we do not know the relative position in the individual life cycle for employees who have not exited the firm. We therefore cannot standardize their tenure.

To address these problems, we employed a matched-pairing approach. We paired each departed employee with exactly one non-departed employee who joined the firm in the same month. Where there was more than one pairing candidate, we randomly assigned a non-departed employee. Departed employees who could not be matched and unmatched non-departed employees were thus excluded from this analysis. The matched-pairing approach allowed us to observe pairs of individuals in the same time frame and compare non-departed and departed employees' enculturation paths. By matching on month of entry, we also addressed unobserved heterogeneity that is time-related. In particular, we included matched-pair fixed effects in our model. We estimated the following model:

$$Y_{it} = \gamma_1 T_{it} + \gamma_2 V_i T_{it} + \gamma_3 IV_i T_{it} + X_{it}\beta + \alpha_p + \epsilon_{it} \tag{SM3.4}$$

where Y_{it} is cultural fit for individual i observed at time t, T_{it} are standardized tenure parameters, V_i and IV_i are dummies for voluntary and involuntary departure, respectively, X_{it} are individual-time observed predictors, β and γ are estimated coefficients, α_p is the unobserved time-invariant matched-pair fixed effect, and ϵ_{it} is the error term. Because departure status is a fixed individual attribute, this specification does not allow for individual fixed-effects; however, it enables comparisons across departure statuses.

The estimated coefficients are reported in Table SM3.4. We estimated three models. In the first, we assumed linear tenure effects. The estimated coefficients imply a positive effect of tenure on cultural fit for non-departed individuals, a negative effect for involuntarily departed, and an effect insignificantly different from 0 for voluntarily departed. The individual fixed-effects model reported in Table SM3.3 suggests that this flat linear effect is a result of a curvilinear relationship between tenure and fit for voluntarily departed. Indeed, in Model 2 in Table SM3.4 we included a square term for tenure and interacted it with voluntary departure. This interaction term is significant. In Model 3, we added the square terms for all departure statuses. The interactions are only significant for voluntary departures. The marginal effects illustrated in Panel B of Figure 4 correspond to this model.

Finally, we sought to investigate whether these enculturation trajectories are tenure invariant beyond individual differences in enculturation tempo. To do so we replicate Model 3 in Table SM3.4 but use tenure in months, rather than standardized tenure, as the main covariate. Because we no longer need to standardize tenure, we do away with the matched-pairs model. This allows us to include all individuals in the model. We include month fixed-effects to account for period-dependent fluctuations in cultural dynamics, and limit our sample to three years of tenure, so as to prevent biasing the results by non-departed individuals (who are significantly over-represented beyond 3 years of tenure). Once again, we used non-departed employees as the omitted category, and added interactions for exit type. We estimated the following model:

$$Y_{it} = \gamma_1 T_{it} + \gamma_2 V_i T_{it} + \gamma_3 IV_i T_{it} + X_{it}\beta + \alpha_t + \epsilon_{it} \tag{SM3.5}$$

where Y_{it} is cultural fit for individual i observed at time t, T_{it} are tenure parameters, V_i and IV_i are dummies for voluntary and involuntary departure, respectively, X_{it} are individual-

Table SM3.4: Fixed-Effects Models of Cultural Fit

	Matched-Pair Fixed-Effects			Period Fixed-Effects
	(1)	(2)	(3)	(4)
Tenure	0.249***	0.558*	0.436	0.015***
	(3.82)	(2.34)	(1.59)	(3.48)
Tenure^2			-0.172	-0.000
			(-0.70)	(-0.56)
Voluntary	-0.050	-0.301*	-0.324**	-0.430***
	(-0.62)	(-2.46)	(-2.59)	(-5.45)
Voluntary x Tenure	-0.269*	1.143*	1.270*	0.046***
	(-2.12)	(2.25)	(2.41)	(3.74)
Voluntary x Tenure^2		-1.637***	-1.470**	-0.001**
		(-3.93)	(-3.05)	(-3.13)
Involuntary	-0.076	-0.071	-0.164	-0.505***
	(-0.92)	(-0.87)	(-1.24)	(-6.92)
Invoulntary x Tenure	-0.666***	-0.665***	-0.201	0.018
	(-5.35)	(-5.36)	(-0.38)	(1.56)
Involuntary x Tenure^2			-0.425	-0.001
			(-0.90)	(-1.71)
Female	-0.198***	-0.200***	-0.199***	0.045*
	(-3.99)	(-4.05)	(-4.01)	(2.18)
Manager	0.217***	0.219***	0.218***	0.621***
	(3.83)	(3.86)	(3.84)	(20.94)
Age	0.014	0.012	0.012	0.048***
	(0.88)	(0.77)	(0.77)	(6.24)
Age^2	-0.000	-0.000	-0.000	-0.001***
	(-0.83)	(-0.73)	(-0.72)	(-6.22)
Intercept	-0.098	-0.121	-0.094	-1.001***
	(-0.33)	(-0.40)	(-0.31)	(-6.81)
Department Controls	Yes	Yes	Yes	Yes
Matched-Pairs FE	Yes	Yes	Yes	No
Period FE	No	No	No	Yes
N	3052	3052	3052	9044
R^2	0.522	0.525	0.525	0.202

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

time observed predictors, β and γ are estimated coefficients, α_t is the unobserved monthly fixed effect, and ϵ_{it} is the error term. The estimated coefficients are reported in Model 4 in Table SM3.4. The marginal effects illustrated in Panel A of Figure 4 correspond to this model.

Early Enculturation

The results in Table SM3.4 Model 4, as well as rhe trends depicted in Figure 4, suggest that early enculturation is particularly important. Is there a time window during which, if cultural adaptation is unsuccessful, the likelihood of failed integration increases? We explore this by measuring cultural fit during an individual's first six months at the firm. As Figure 3 demonstrates, individuals on average reach mean cultural fit by the end of their first year at the firm; the increase is particularly steep during the first six months immediately post entry. This is consistent with previous research on enculturation (Bauer et al. 2007). Six months provide enough data points to estimate a trend. They exclude a number of individuals who leave the firm less than six months after joining it, but not enough to undermine statistical power. We estimate an individual's rate of enculturation by fitting the following simple linear model $Y_{it} = \beta_0 + \beta_1 t + \epsilon_{it}$, where Y_{it} is cultural fit, t is month, and β_1, the slope of the fitted line is the rate of enculturation. A rate of 1 implies an increase of one standard deviation in cultural fit per month (though this is an especially high rate; the interquartile range is [-.095 .134]). Note that because some individuals depart the organization before six months, especially involuntarily, and because others have missing initial cultural fit values at time of entry due to early low activity, the number of exit events reported in Table 1 is lower than the overall number of exits observed in the dataset.

We next estimated a Cox proportional hazard model as in eq. SM3.1, where predictors include enculturation rate during the first six months and initial cultural fit (observed during the individual's first month at the firm). We estimated two models, one for involuntary and one for voluntary exit, as reported in Table 1. Enculturation rate strongly predicts involuntary exit but not voluntary exit: a one standard deviation increase in cultural fit per month decreases the hazard ratio of involuntary exit by 91% (or 35.1 percentage points for an individual at the 95^{th} percentile of enculturation). In Figure 5 we plot cumulative hazard functions for involuntary exit for different newcomers. We defined newcomers with high cultural fit as those at the 75^{th} percentile of initial fit and those with low fit as those entering at the 25^{th} percentile of fit. We defined highly adaptable employees as those at the 95^{th} percentile of enculturation and those highly inadaptable at the 5^{th} percentile. We then calculated hazard functions for median newcomers (at median levels of initial fit and adaptation), high initial fit-adaptable (at 75^{th} percentile of initial fit and 95^{th} percentile of adaptation), high initial fit-inadaptable (at 75^{th} percentile of initial fit and 5^{th} percentile of adaptation), low initial fit-adaptable (at 25^{th} percentile of initial fit and 95^{th} percentile of adaptation) and low initial fit-inadaptable (at 25^{th} percentile of initial fit and 5^{th} percentile of adaptation). We are particularly interested in low initial fit-adaptable newcomers, who, despite entering the organization with low initial cultural fit, are able to adapt culturally at a fast rate and overcome the negative implications of initial low fit. These individuals' hazard of involuntary exit is lower than that of the median newcomer or that of newcomers entering the organization with high fit but an exceptionally low cultural adaptation rate.

References

Bauer, Talya N., Todd Bodner, Berrin Erdogan, Donald M. Truxillo, Jennifer S. Tucker. 2007. Newcomer adjustment during organizational socialization: A meta-analytic review of antecedents, outcomes, and methods. *Journal of Applied Psychology* **92**(3) 707–721.

Cox, David R. 1972. Regression models and life tables. *Journal of the Royal Statistical Society: Series B* **34** 187–202.

Greene, William H. 2012. *Econometric Analysis*. 7th ed. Prentice Hall, Upper Saddle River, NJ.

Pennebaker, James W., Cindy K. Chung, Molly Ireland, Amy Gonzalez, Roger J. Booth. 2007. The development and psychometric properties of liwc2007. Tech. rep., LIWC.net.