THE UNIVERSITY OF CHICAGO

CULTURAL ASSIMILATION AS A HUMAN CAPITAL FORMATION PROCESS: THEORY AND EMPIRICAL EVIDENCE

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF ECONOMICS

BY
JAMES VINCENT MARRONE

CHICAGO, ILLINOIS
JUNE 2017

for David,

with much admiration and love for assimilating so easily and willingly

to life among economists

. . . the image of a constant human nature independent of time, place, and circumstance, of studies and professions, transient fashions and temporary opinions, may be an illusion, that what man is may be so entangled with where he is, who he is, and what he believes that it is inseparable from them.

–Clifford Geertz, *The Interpretation of Cultures*, p. 35 (Basic Books, 1973)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Most importantly, thanks to my committee members for their unconditional support and interest in what was originally just a germ of an idea: Steven Durlauf, Casey Mulligan, and Kevin Murphy. To my study group (John Eric, Karl, Ali, Fabrice, and Pat): thanks for six great years together in Chicago, and hopefully many more as friends and colleagues around the world.

Thanks also to Ed Lazear for particularly timely discussions on how to make these ideas successful in the outside world. Finally, seminar audiences at the UChicago Applications Workshop, University of Wisconsin-Madison, WEAI Graduate Student Workshop, RAND Corporation, and Clemson University provided helpful comments on various chapters of this dissertation.

ABSTRACT

Unlike economic assimilation – the process of immigrants' wages converging to those of natives – cultural assimilation – the process whereby people adopt specific beliefs, languages, or modes of daily life to conform to those around them – has not been studied in depth using economic tools. This thesis examines cultural assimilation and cultural change as processes resulting from human capital investments. Unlike past studies, where assimilation is viewed as a passive process or as serving to signal identity in a static world, a human capital approach to assimilation emphasizes the productive aspects of culture and nests assimilation into a framework of supply and demand, with intertemporal constraints and forward-looking actors. I focus on three aspects of assimilation that have not been fully examined: the theory of cultural change, empirical evidence for different qualitative types of cultural change, and the proper measurement of cultural human capital.

In the first chapter, I develop a learning-by-doing theory of cultural assimilation, in which an individual's cultural identity is determined by past investments in culture-specific social capital. The model incorporates several empirically relevant aspects of assimilation that have been difficult to explain in past models. The mechanism in my model yields two possible qualitative outcomes of the assimilation process, depending on the intertemporal complementarities of investment. The first, in which individuals become more homogeneous over time, has been used to explain immigrants' wage assimilation. The second, in which individuals become more heterogeneous as some assimilate and some do not, has vastly different implications for immigration and integration policies. As a purely theoretical exercise, the chapter offers a new model of group formation: rather than emerging from individuals forming and erasing costly bonds in a multi-period game, social networks can form endogenously via investment in productive capital.

In the second chapter, I use detailed datasets from three immigrant destination countries to evaluate whether assimilation patterns adhere to the predictions of the first or second type of process from Chapter 1. The data provide a variety of evidence from linguistic,

identity, and religiosity measures that cultural assimilation (in contrast to economic assimilation) most closely resembles the second type of process. The regressions typically used on cultural survey data are ordered probit or logistic regressions, or studies will collapse ordinal variables into binary data. A major point of this chapter is that such regressions may be statistically inappropriate, and at the very least can miss important information about the second moment of the distribution – for example, the fact that there is a long tail of individuals who assimilate very slowly, and who are possibly the most important targets of policy interventions.

In Chapter 3, I construct new measures of linguistic human capital. Using the same data from Chapter 2, I show that different measures of linguistic skills (such as speaking, writing, and reading) measure different dimensions of language fluency. Thus, a proper measure of linguistic skill must incorporate multiple dimensions of evidence. I estimate a continuous latent trait for linguistic skill using a set of 11 survey questions. The methodology comes from the psychometrics literature, designed specifically for analyzing ordinal test data of the sort used to measure language skills. The measures control for systematic differences in the ways different groups answer questions (Differential Item Functioning, or DIF) across groups. I use different latent factor constructs along with standard measures of past studies (binary variables or simple averages of reading/speaking/writing/understanding) to re-estimate the determinants of language skills and the wage premium for learning a language. I show that the results are not sensitive to the exact form of the latent factor, but they are quite different from the results using only binary or discrete variables. In particular, past studies have misestimated the wage premium for language skill; I find that there is a premium for speaking, and an additional premium (of almost the same size) for reading/writing in addition to speaking.

CHAPTER 1

A RATIONAL CHOICE THEORY OF ENDOGENOUS GROUP FORMATION

1.1 Introduction

Cultural assimilation – the process of becoming more like a specific reference group along some quantifiable margin – has become a major political concern. For example, a 2016 report notes that the majority of Europeans say growing diversity makes their country a worse place to live, and believe that Muslims do not assimilate (Pew, 2016). Assimilation is also an economic issue: social science offers mounting evidence that assimilation promotes economic and civic well-being (see Vigdor, 2015, for a discussion). A burgeoning literature applies standard economic theory to the question of how individuals' identities form and why they may be relevant in explaining outcomes such as labor force participation, wages, marriage, or fertility. In general, extant models of identity formation treat identity as either the equilibrium outcome of a game in which players have explicit social preferences, or as the result of human capital investments in a simple (often static) framework with culture-specific (often binary) capital stocks. The former approach has been used most fruitfully by Bisin & Verdier (2000a) and Akerlof & Kranton (2010) to explain the signaling value of identity, while the latter has been leveraged by Borjas (2000) and Lazear (1999) to explain wage assimilation and the formation of ethnic enclaves.

But several important aspects of cultural assimilation - and cultural change more generally - remain unmodeled and poorly understood. Importantly, theory has failed to account for the entire distribution of assimilation outcomes. While much research has shown that the average immigrant assimilates along several margins (including cultural margins, such as linguistically), people generally assimilate at different rates, and many do not assimilate at all. These facts imply that the variance of outcomes is just as important as the mean, and that the time-dependent nature of the process must be modeled explicitly. In addition, previous models have ignored the dual nature of cultural assimilation: people simultaneously un-assimilate from their native culture as they assimilate to a new one. One test of a satisfactory theory of cultural change is that it should be able to explain both aspects of the process via the same mechanism; thus far, theories have failed this test. Finally, past models have failed to account for the apparent durability of cultural practices: history offers numerous examples of cultural practices which were banned or suppressed but which resurged afterwards, sometimes after decades of dormancy. This strongly suggests that some expressions of identity, rather than being merely signals of group membership that can be turned on or off, are symbolic manifestations of productive, durable human capital stocks.

In this chapter I model an economic mechanism behind cultural assimilation in order to capture these aspects of cultural change and to derive empirically relevant predictions. In the model, individuals divide their time endowments between two culture-specific stocks of capital ("local" versus "foreign"). In contrast with the conceptual precedents of Dustmann (1999) and Borjas (2000), who model a single stock that can be transferred between cultures, the two-stock setting fully endogenizes all costs and generates a variety of long-run outcomes. And in contrast to the static model of Lazear (1999), my model's dynamic setting allows me to distinguish different qualitative trajectories. At the heart of the model is the ambiguous effect of assimilation on the incentive to assimilate further, which can yield different qualitative outcomes with very different implications for long-run human capital stocks. For example, consider language fluency: on the one hand, there may be diminishing returns to increased fluency, but on the other hand there may also be diminishing marginal costs. Whether an immigrant becomes fluent in a language depends on these relative costs and benefits.

I derive conditions under which the benefits outweigh the costs, implying individuals either assimilate fully or resist assimilation altogether. If this does not occur, we would expect only partial investment in either cultural stock, yielding a multi-cultural "melting pot" in which people become more alike over time. But if the benefits do outweigh the costs, we would expect increasing heterogeneity, with similar individuals becoming very different over time as one assimilates and one does not; essentially, people specialize in one culture or identity at the expense of the other. In the extreme case, enclaves or oppositional identities could form - for example, when some immigrants become more intensely religious while others become more secular than before. This produces a bifurcation in investment trajectories; the mathematical conditions yielding this phenomena characterize this type of assimilation process as an addictive process in the sense of Becker & Murphy (1988).

The two states of the world have very different implications for assimilation policy. In the case of increasing homogeneity, cultural investments wane over time, and any policy to increase a cultural stock will crowd out future investments without affecting individuals' long-run outcomes. But in the case with increasing heterogeneity and addictive culture, investments are complementary over time and policies could actually amplify later investments. What's more, the individual's optimal investment function can be discontinuous: that is, there will be people at or near a margin dividing assimilators from non-assimilators. These people are especially susceptible to policies that push them to the other side of the margin, resulting in entirely different trajectories and having great implications for their children's life trajectories as well.

Overall, this paper seeks to illuminate the tradeoffs inherent in assimilation, to understand who assimilates and why, and to begin understanding an economic mechanism behind cultural change. Predicting individual assimilation outcomes is vital for crafting effective integration policy, as well as for understanding intergenerational mobility and economic impacts of cultural change. As noted above, the model in this paper incorporates several important aspects of assimilation in order to provide empirically relevant predictions. Those motivating features of assimilation are discussed in Section 1.2 in the context of previous studies. Section 1.3 presents the basic model and its predictions using linguistic assimilation as a concrete example. Section 1.4 discusses historical, anthropological, and sociological examples relevant to the model, and Section 1.5 outlines directions for future research.

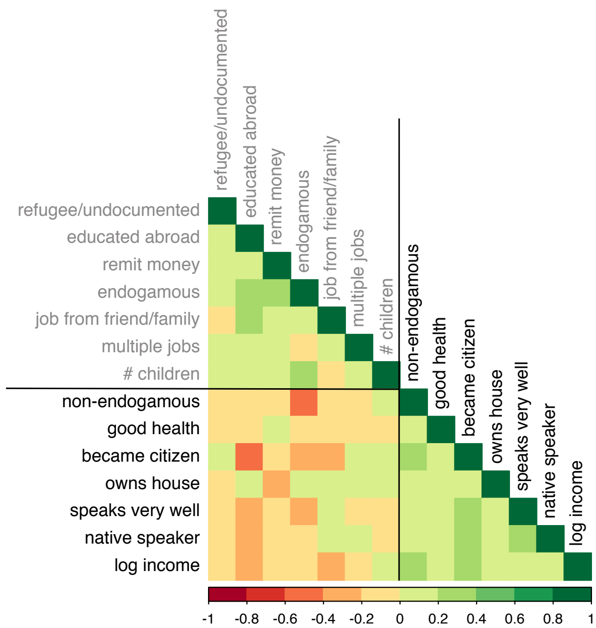
1.2 Motivating Facts and Features of Assimilation

1.2.1 Culture is an Abstract Stock

The theory and empirics together imply that a single underlying mechanism can unite many concepts currently grouped under the umbrella of "assimilation." Past studies have shown similar patterns of convergence across a wide variety of outcomes. Starting with the work of Chiswick (1978) and Borjas (1985), the large economic literature on wage assimilation highlights immigrants' slow wage convergence relative to natives. Blau *et al.* (2011) and Fernández (2013) study convergence in female labor force participation. Abramitzky *et al.* (2014) document the tendency of immigrant families to eventually adopt American names for their children, leaving a 10% wage gap between older children with foreign names, and younger children with American ones. Beliefs and norms can also converge over time, and not necessarily in beneficial ways. For example, Kliewer & Ward (1988) and Kliewer (1991) provide evidence that converging social norms or ideas of right and wrong may lead to observed convergence in suicide rates between immigrants and natives. Relatedly, Pavlish *et al.* (2010) note that Somali women often face frustration and diminished quality of care in their medical visits, insofar as their own beliefs and expectations have not converged to the western, biological model of health with its associated treatments and doctor/patient interactions.

Figure 1.1 broadly summarizes these collective findings while defending one overall premise of the model: empirical indicators of assimilation tend to be positively correlated, and hence it makes sense to think of a single abstract "cultural stock" behind the assimilation process. The figure aggregates data from all three surveys used in this paper, collectively recording 32,684 individuals from all over the world who moved to either the USA, Spain, or France. The variables labeled in black are most commonly used as measures of social or economic assimilation: intermarriage, level of income, language fluency, homeownership, naturalization, and overall health. Other variables, labeled in gray, indicate lack of assimilation: religiosity, family size, having multiple jobs, or being a refugee. Despite many differences across sending and receiving countries, the two sets of variables show positive within-group but negative across-group correlations, implying that individuals tend to assimilate along multiple margins simultaneously, or not.[1] This supports the modeling framework that assimilation is a holistic process with a single underlying mechanism.

Figure 1.1: Correlations Between Social and Economic Outcomes



Note: The figure shows the pairwise correlations between various outcomes, using aggregated data from the NIS, ENI, and TeO surveys. Variables implying assimilation along some margin are labeled in black; variables implying the opposite are labeled in gray.

Sociologists have spent much time classifying the connections between different margins of assimilation. The broader literature classifies two different "flavors" or dimensions of assimilation. "Structural assimilation" consists of those dimensions that affect socioeconomic

1. The major exceptions are language skills ("speaks very well" and "is a native speaker"), which are positively correlated with using friend/family social networks to find a job, and having multiple jobs. However, this is entirely attributable to the presence of many Latin American immigrants in Spain (48% of the survey), who share a native language with Spaniards but who otherwise share more characteristics with other immigrants. If Latin American immigrants to Spain are dropped from the sample, then the positive correlations turn negative.

outcomes, while "acculturation" encapsulates those that describe cultural patterns.[2] The two are certainly not the same, as White & Glick (2009, p. 31) note in their survey of the literature: "One may observe parity in income or education levels, yet also a society in which ethnic differences are highly manifest in other domains, such as marriage, friendship ties, or neighboring." Yet scholars still debate the degree to which structural assimilation may differ from acculturation, and which margins of assimilation fall in which category. It is not entirely obvious, for instance, that beliefs about gender roles should be grouped under acculturation, since those beliefs can impact female labor force participation and thereby affect income (a pattern shown elegantly by Fernández, 2013). By eschewing the structural/acculturation distinction in favor of a single holistic assimilation process, this paper provides a single assimilation mechanism, which could in turn lead us to think about what causes differences between various manifestations of the process.

Of course, the correlations in Figure 1.1 are not strictly the same across every margin for every group, which is why sociologists have developed theories of "segmented assimilation." The sociological classification of trajectories refers to whether or not a person or group will assimilate. The first type of trajectory, straight-line assimilation, yields convergence to natives over time, both within and across generations. But means often converge even though some individuals assimilate at different rates; and even at the group level, some groups fall further behind the majority over time. To account for the failure of straight-line assimilation theory in certain cases, researchers developed "segmented assimilation" models (Gans, 1992; Portes & Zhou, 1993). The original theory emphasized barriers: discrimination and legal impediments can prevent entire groups from integrating into society. Barriers can certainly explain the downward paths observed for several groups, but segmented assimilation theory remains unsatisfactory without a well-defined model. The concept might best be

2. In his foundational work, Gordon (1964) actually outlined seven types of assimilation: acculturation, structural, amalgamation (intermarriage), identificational (sense of group identity), attitude receptional (absence of prejudice from other groups), behavior receptional (absence of discrimination), and civic. These more-specific categories pose the same problems as the structural/acculturation dichotomy.

characterized as saying that minority groups do not have to adapt to the majority. In fact, some of the original theorists summarize their own theory as follows: Segmented assimilation is "a process... where the outcomes differ across immigrant minorities and where the rapid integration and acceptance into the American mainstream represent just one possible alternative" (Rumbaut & Portes, 2001, p. 6). For an economist looking to formulate empirical predictions, it is difficult to proceed from such a point. Nevertheless, the segmented assimilation concepts makes the important point that some groups do not neatly converge to the majority. This point has been occluded in the economics literature, but allowing for non-assimilation is one virtue of my model.

1.2.2 Assimilation is a Dual Process

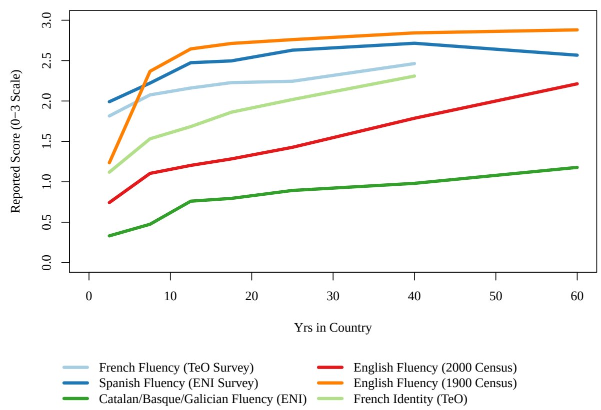
Another reason to model assimilation as a human capital process is because assimilating to one group often means un-assimilating from one's original group. The positive and negative correlations in Figure 1.1 imply this in part, and it necessarily holds for certain outcomes, such as religiosity (one cannot simultaneously get both more and less religious). But it is not obvious that it must be true for language or self-identity (people can be bilingual or hold multiple self-identities). Nevertheless, Figure 1.2 provides empirical evidence for the phenomenon. As immigrants assimilate in language and identity (Panel a), they also un-assimilate from their native language and birthplace (Panel b). By construction, this same pattern occurs in the model, which treats assimilation and de-assimilation as two sides of the same coin, part of a human capital investment-and-depreciation process.[3] While un-assimilation has often been overlooked in past studies, it is potentially a major cost of assimilating to a new culture. Losing one's native culture, with both the memories of a continuous identity as well as the ingrained habits that form one's behaviors, is plausibly a major psychological cost and could be a contributing factor to the slow nature of cultural

3. Psychologists and linguists have developed a substantial literature on how and why people lose fluency in their mother tongues. Labeled "first language attrition," the process may have a biological basis. See Schmid *et al.* (2004); Schmid (2011) for overviews of the relevant studies.

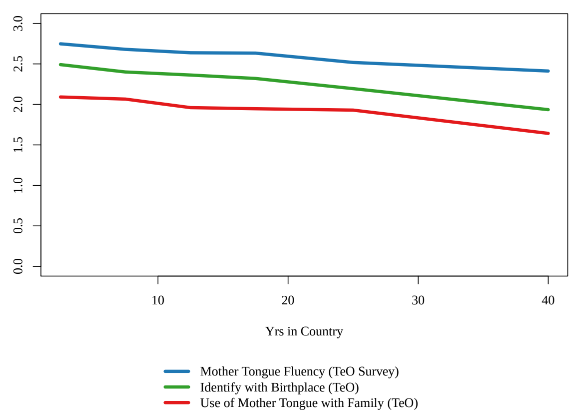
change.

Figure 1.2: Assimilation and Un-Assimilation Paths

(a) Assimilation Indicators



(b) Unassimilation Indicators



Note: Figures show average reported characteristics, on 0-3 scale, for cohorts based on tenure. Survey sources are noted in parentheses.

To capture the dual nature of the assimilation process, it is important to incorporate two cultural stocks. A noteworthy example is Lazear (1999) (expanded by Kónya (2007)), who constructs a static model in which human capital is a binary variable representing language fluency. The outcome of interest is the equilibrium level of linguistic diversity, a topic of special interest to the growth and development literature, as in Alesina & Spolaore (1997);

Alesina *et al.* (2003). Diversity could be interpreted as the extent of homogenization of groups, which is akin to assimilation. However, Lazear does not allow for people to "forget" their native language; his is, essentially, a model of bilingualism.

Closer to my model are the game-theoretic treatments of Bisin & Verdier (2000a,b, 2001), which describe another possible mechanism behind cultural change; specifically, a socialization technology that operates both within the family and in the larger community. Following on these papers, Bisin *et al.* (2004, 2011, 2016) use similar models to examine religious intermarriage and immigrants' ethnic identities. Like my human capital investment model, the socialization technology hinges on the ambiguous relationship between costs and benefits of assimilation. For example, in Bisin *et al.* (2016), immigrants' marriage decisions are determined by the costs of interacting with natives. But those costs may either rise or fall with the proportion of other immigrants in one's neighborhood, and so the propensity to intermarry could rise or fall as well (see Proposition 1 in particular). This is a stylized formulation of the same relationship between costs and benefits in my model, which determines whether or not the bifurcated, bimodal distribution of outcomes will eventually form in the long run. But like Lazear (1999), these models assume exogenous costs, and the endogenous duality of the assimilation/un-assimilation relationship is un-modeled.

1.2.3 Assimilation is Time-Dependent

Figure 1.1 also clarifies the time-dependent nature of assimilation, since outcomes are positively or negatively correlated with time since migration. The pattern is well-established, particularly for language fluency, and has been studied at least since Chiswick (1978). But beyond language fluency, several other behaviors are linked to time since migration. These include self-reported feelings of identity, but also behaviors such as trust in civic institutions, participation in civic organizations, religiosity, and others. In this sense, then, even when behaviors serve a signaling purpose (such as the pioneering model in Akerlof & Kranton (2000)), the value or meaning of that signal may be tied to an underlying latent stock of cultural identity.

My model operates in continuous time, which allows me to distinguish trajectories from long-run outcomes. Trajectories are determined by an individual's initial characteristics; in the sociological literature, Stolzenberg (1990) coins the term "conditional assimilation" to describe this concept. One virtue of continuous time is that we can look at both the rate and the direction of assimilation. The conceptual precedents here are Glaeser *et al.* (2002) and Grin (1992), who also present fully dynamic models. Glaeser *et al.* (2002) seek to explain the accumulation of social capital such as trust, and so they use only one capital stock. This generates different predictions from my model; in particular, the mathematical conditions for addiction are not necessary to generate any specific outcomes. But like this paper, they do use survey data (in their case, the General Social Survey) to provide empirical evidence for their predictions. Grin (1992) is interested in the sustainability of minority languages and is closest to my model. Grin's stocks capture population-wide averages and reflect the level of vitality of a language, rather than individual capital stocks within a population. In addition, my model makes fewer parametric assumptions.

Another major precedent is Borjas (2000), who presents a two-period human capital model to explain why the wage gap between immigrants and natives closes over time. Borjas uses a single-stock model (also in Borjas (2014) and Borjas (2015)) based on Ben-Porath (1967), which predicts that people should make human capital investments early in life, and reap the benefits later. This qualitative outcomes matches the increasing-homogeneity case of my model, and in that sense my model may be seen as a generalization of these prior formulations.

1.2.4 Human Capital Investments Can Increase or Decrease Over Time

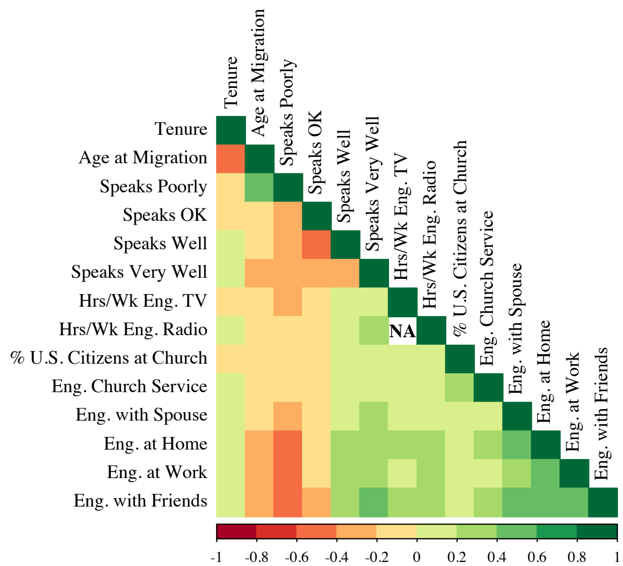
However, the Ben-Porath model fails to explain all aspects of assimilation. In particular, it predicts that investment today displaces investment tomorrow. In addition, it predicts that investments decrease over time. This may be true for aspects of human capital such as education, but it fails for, say, language use: immigrants eventually use their second language in more and more spheres of daily life as they improve their fluency. This may seem utterly obvious, but it is also clear in the data. Figure 1.3 illustrates the correlation between language fluency and language use, in the New Immigrant Survey (USA) and the Trajectories and Origins survey (France). The qualitative patterns are the same for both surveys. It is immediately apparent that speaking well or very well positively correlates with using the language and listening to English media. These variables are also positively correlated with tenure and negatively correlated with age at migration. On other hand, speaking poorly or merely OK is correlated in the opposite direction.[4] This is evidence that human capital investments build over time, and that people eventually "specialize" in both knowing and speaking a language.

Figure 1.3 asserts that the Ben-Porath framework is insufficient to explain every form of assimilation or human capital investment process. My model nests the Ben-Porath-style investment process in the same setting as the increasing-investment process; the only difference is the strength of intertemporal complementarities.

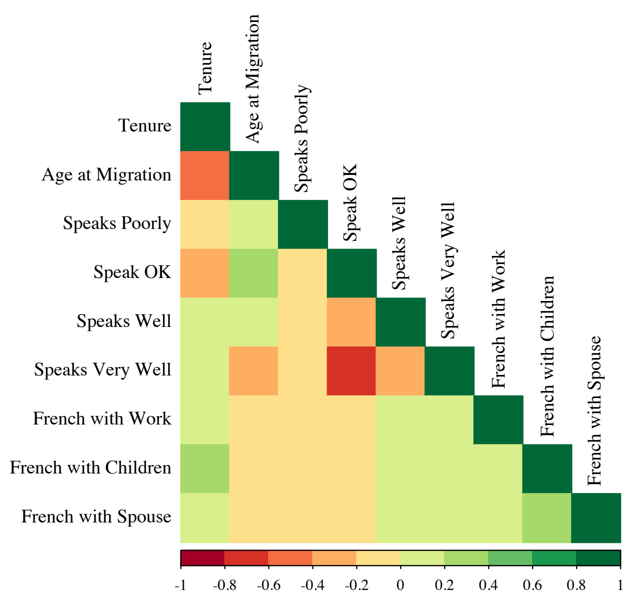
4. The missing correlation between radio and TV in the NIS panel is because each respondent was asked about radio or TV, but not both.

Figure 1.3: Correlation Between Language Use and Fluency.

(a) NIS Survey



(b) TeO Survey



Note: Figures show pairwise correlations between levels of language fluency and use of that language in daily life. The relevant language is English for the NIS survey and French for the TeO survey.

1.3 Model

To model cultural assimilation, I adapt well-known learning-by-doing models in which the productivity of time in an activity depends on the stock of experience in that activity. Here, the activity is the production of generic consumption goods using time investments and of two culture-specific stocks. In order to allow for investments to have both positive and negative intertemporal complementarities, the model makes use of the framework developed in Stigler & Becker (1977) and Becker & Murphy (1988), generalized to include more than one good as in Iannaccone (1986) and Dockner & Feichtinger (1993). In these models, the goods were assumed be addictive, and the goal was to explain how addictiveness (with consumption investments increasing as the underlying stock increases) may arise from standard economic assumptions (Becker *et al.* , 1994; Dragone *et al.* , forthcoming). It is important to note that in this framework, addiction arises organically as a necessary but not sufficient condition for complete specialization in or assimilation to one of the two cultural stocks. The other possibility, in which culture is not addictive, has equally plausible interpretations and empirical implications; thus, the model clarifies and distinguishes two possible varieties of human capital investment that can map to assimilation trajectories. In this way, the model is related to previous studies of religion (Iannaccone, 1984) and the arts (Castiglione & Infante, 2015).

Throughout the exposition, language is used as a concrete example so that the stock intuitively represents linguistic fluency, but cultural capital could also be reflected in self-identity, religiosity, or numerous other characteristics. The individual divides a fixed amount of time each period between speaking each language, which in turn allows the individual to produce all consumption goods.[5] Thus, unlike several previous models of addiction, there is no numeraire good: the optimal investment path could result in a corner solution, in which the individual speaks only one language. Such a long-run outcome can be easily interpreted

5. The utility function can be interpreted as an indirect utility function in which decisions over labor supply and consumption have been maximized out.

in this particular setting as monolingualism, with an interior steady state corresponding to bilingualism. But the possibility of corner solutions distinguishes this model from its predecessors, meaning that we cannot rely on previous assertions by Iannaccone (1986) and Dockner & Feichtinger (1993) that, under certain conditions, the two cultural stocks should monotonically approach separate steady states.

For the baseline model, suppose an immigrant arrives in a new country at time 0 with time horizon T. At each time t the immigrant may divide one unit of time between two languages, the local language $\mathcal{L}$ and their mother tongue $\mathcal{M}$. There is only one consumption good, which can be produced with either $\mathcal{L}$ or $\mathcal{M}$. Marginal productivity of time spent speaking each language also depends on the immigrant's experience speaking that language, given by stocks L and M.

Following Becker & Murphy (1988), utility is a generic function of current and past consumption. The stocks of $\mathcal{L}$ and $\mathcal{M}$ are given by $L(t)$ and $M(t)$, and time investments in each stock are $\ell(t)$ and $m(t)$. The parameter vector α characterizes the utility function. The individual maximizes utility over remaining horizon T, as given by equation 1.1.[6]

$$U(0) = \max_{\{\ell(t)\}} \int_0^T e^{-\rho t} u\left(\ell(t), L(t), m(t), M(t); \alpha\right) dt \tag{1.1}$$

Assume u is quasi-concave, so that preferences are convex. In addition, assume positive cross-partials between each corresponding time input and stock: in other words, being more fluent in the language makes a person more productive. Finally, language stock should not affect productivity of speaking another language. These assumptions are summarized in

6. The lifespan T may also be interpreted as expected tenure. A generalized version of this model could incorporate continuation values at T, at which time the individual would return to their native country. In this way the model can be related to the literature on temporary migration. See Dustmann (2003); Dustmann & Görlach (2016).

equation 1.2.

$$u_i > 0 \,\forall\, i = \ell, L, m, M \tag{1.2a}$$

$$u_{ii} < 0 \,\forall\, i \tag{1.2b}$$

$$u_{\ell L} > 0 \tag{1.2c}$$

$$u_{mM} > 0 \tag{1.2d}$$

$$u_{\ell m} = 0 \tag{1.2e}$$

$$u_{\ell M} = 0 \tag{1.2f}$$

$$u_{mL} = 0 \tag{1.2g}$$

$$u_{LM} = 0 \tag{1.2h}$$

Assumptions 1.2e-1.2g yield a more tractable optimal control problem but are not strictly necessary for any of the results. In the rhetorical convention of past studies, they state that time use is commodity-specific. This is not realistic for some physical consumption goods, where the activities themselves might be complements, as in Dragone *et al.* (forthcoming). However, in this setting the assumptions are motivated by realism as well as simplicity. Assumption 1.2e implies time spent speaking one language has no effect on the marginal productivity of time in another language. If this were false, languages would be substitutes or complements and bilingualism would be inherently built into the utility function, rather than the end product of language use. Assumptions 1.2f-1.2h rule out spillover effects between languages. These are realistic insofar as knowing one language does not make someone more fluent in another. This may not be true if languages are mutually intelligible, and the assumptions also preclude possible hybrid forms of communication such as pidgin English or Spanglish. From an identity viewpoint, the assumptions preclude mixed-identity labels such as "Japanese-American." These aspects of language or identity evolution may yield fruitful extensions for future consideration.

Assume the stocks depend on the period investments and a depreciation parameter δ, as defined in equation 1.3.

$$\dot{L}(t) = \ell(t) - \delta L(t) \tag{1.3a}$$

$$\dot{M}(t) = m(t) - \delta M(t) \tag{1.3b}$$

Again, some simplifying assumptions should be justified by reality. First, equation 1.3 implies commodity-specific consumption capital: fluency is a function of time spent using that language, and that language alone. Linguistically, this assumption rules out spillover effects between languages, or instances of mutual intelligibility. Again, these concepts may offer fruitful future extensions to the model, but they are ignored for now.

At this point, the system defines two stocks, two controls, two equations of motion, and a period-by-period budget constraint. The dynamic system appears to yield a standard optimal control problem; compare, for example, Dragone *et al.* (forthcoming, §3.1), who additionally include a numeraire. A few more points, however, add both realism for this particular setting and a simpler re-formulation of the Hamiltonian. Since the period-by-period time constraint is $\ell(t) + m(t) = 1$, we can reduce the problem to maximizing over one parameter $\ell(t)$, the fraction of time spent speaking $\mathcal{L}$. The final assumption sets initial conditions on the total stock to equal $\frac{1}{\delta}$. This ensures that the stocks will always sum to $\frac{1}{\delta}$, so that changes in stocks reflect pure changes in investment, rather than growth in total stock over time. Essentially, growth of the total stock would be like an income effect that complicates the intuition without adding any generality to actual mechanism underlying assimilation. Combining both the stock and time constraints, equation 1.4 describes the joint evolution of the stocks.

$$L(0) + M(0) = \frac{1}{\delta} \tag{1.4a}$$

$$\Rightarrow L(t) + M(t) = \frac{1}{\delta} \ \forall t \tag{1.4b}$$

$$\dot{M}(t) = -\dot{L}(t) \tag{1.4c}$$

Equation 1.4 provides an easy definition of “assimilation:” a person’s relative stock level at any given time measures the degree of their assimilation. Since the total language fluency is constant, the relative allocation reflects how much they have spoken that language, on average, in the past.

Combining all of the constraints and writing M and m in terms of L and ℓ, the two-stock, two-good model reduces to a dynamic system with one state and one control. The resulting Hamiltonian is in equation 1.5 and the first-order conditions are listed in equation 1.6.

$$\mathcal{H} = u\left(\ell, L, 1-\ell, \frac{1}{\delta} - L\right) + \lambda\left(\ell - \delta L\right) \tag{1.5}$$

$$0 = u_\ell - u_m + \lambda \tag{1.6a}$$

$$\dot{\lambda} = (\rho+\delta)\lambda - u_L + u_M \tag{1.6b}$$

Differentiating the first-order conditions and rewriting, the differential equations in equation 1.7 characterize the system in (ℓ, L)-space.

$$\dot{L} = \ell - \delta L \tag{1.7a}$$

$$\dot{\ell} = \frac{\left(u_{\ell L} + u_{mM}\right)\left(\ell - \delta L\right) - \left(u_L - u_M\right) - (\rho+\delta)(u_\ell - u_m)}{-u_{\ell\ell} - u_{mm}} \tag{1.7b}$$

In general, the policy function $\hat{\ell}(L, T)$ can be discontinuous in both L and T (see Figure 1.8 and discussion). Apart from the mathematical specifics, however, the empirically-relevant questions are whether and when $(\ell, L) \to (0, 0)$ or $\left(1, \frac{1}{\delta}\right)$ as $t \to T$.

1.3.1 System Dynamics and Long-Run Outcomes

The question of whether individuals will assimilate in the long run partly depends on whether time investment ℓ increases with the stock L. Even though $u_{\ell L} > 0$, increasing L could raise the shadow price of ℓ; for consumption to increase, the marginal utility must rise by more than the shadow price. By definition this is addiction, as highlighted in Definition 1. The necessary and sufficient condition for this to occur can be derived using the Volterra derivative (see Dockner & Feichtinger, 1991, §4); hence, the condition is treated as a definition (2) and is understood to be equivalent to addiction. This equivalent condition is known as adjacent complementarity and is invoked regularly in the literature. As the name implies, it means that investments today are complementary to investments in the near future.[7] Addiction and adjacent complementarity will be used interchangeably throughout the remainder of the paper.

Definition 1 (Addictive Good, Adjacent Complementarity)**.**
Language is addictive if $sgn\left(\dot{\ell}\right) = sgn\left(\dot{L}\right)$.

Definition 2 (Adjacent Complementarity)**.**
Language displays adjacent complementarity whenever $u_{\ell L} + u_{mM} + \frac{u_{LL}+u_{MM}}{\rho+2\delta} > 0$.

Intuitively, the two definitions are equivalent for the following reason: the formula in Definition 2 is the effect of a marginal increase in stock L on the marginal return to investment ℓ. The first term is the direct, contemporaneous effect. The second term is the future effect, discounted by a factor of $\rho + 2\delta$, one factor of δ from the equation of motion 1.7 plus the net discount factor $\rho + \delta$. Thus, when raising L also raises the overall marginal return to ℓ, ℓ and L should move in the same direction.

In this model, addiction would correspond to a system in which language stocks and investments grow together. This does not, however, imply increasing heterogeneity with trajectories diverging over time. In fact, addiction does not necessarily even imply that all trajectories lead towards a corner – in other words, long-run monolingualism is not equivalent to addiction. We must distinguish monolingualism as a qualitative outcome from the inherent

7. For further discussion of adjacent complementarity, see Becker & Murphy (1988), Dockner & Feichtinger (1993), and Iannaccone (1986). The original discussion is in Ryder & Heal (1973), although it was first made explicit in Boyer (1983). I rewrite the condition in a nonstandard but more intuitive form.

addictiveness of language, and will determine the cases in which they coincide.

To determine whether a system leads to monolingual outcomes, we must consider the steady states. I assume that there is a unique interior steady state, i.e. a unique point $(\ell^*, \delta\ell^*)$ for which $\ell \in (0,1)$. This assumption streamlines the qualitative classification of different systems but in general it need not be true; there could be more or less than one steady state. However, when multiple steady states exist then each steady state will determine the local dynamics immediately nearby. Hence, the more general cases can be described as combinations of the three basic cases explicated below, as illustrated in Section 1. In addition, when there is no interior steady state, long-run monolingualism must result. This is because all trajectories must necessarily lead to a corner, as there can be no stable arm converging to an interior point. This case is also ignored, as the qualitative dynamics are subsumed by the monolingual case discussed below, but Section 2 provides an illustration.

The stability of any steady state is determined by the roots of the Jacobian at that point. Stability requires one negative root; instability requires two positive or complex roots. With a unique steady state, stability or instability will determine the qualitative dynamics of the rest of the system. In fact, we do not even require transversality conditions, because time investments and stocks are bounded in a compact space. Rather, language must be sufficiently addictive that one language eventually "wins out" and all trajectories lead towards a corner, implying full fluency in one language. This means adjacent complementarity is necessary but not sufficient for complete cultural specialization in the long run. Conversely, distant complementarity is sufficient but not necessary for an interior, bilingual long-run outcome. Result 1 summarizes the mathematical conditions.

Result 1 (Conditions for long-run monolingualism)**.**

When the interior steady state is unique:

1. *Long-run monolingualism holds if and only if* $u_{\ell L} + u_{mM} + \frac{u_{LL}+u_{MM}}{\rho+2\delta} > \frac{\delta(\rho+\delta)(u_{\ell\ell}+u_{mm})}{\rho+2\delta}$

2. *Adjacent complementarity is necessary but not sufficient for long-run monolingualism.*

Combining the threshold determining addiction with the threshold determining stability, we obtain three possible cases, as shown in Figure 1.4. Region 2 is of special interest. In this region, cultural stocks are addictive and yet long-run bilingualism results. In other words, both investment ℓ and stock m increase over time but don't converge to a corner. Culture is not addictive enough that individuals eventually choose to specialize in either one or the other. The next section illustrates the qualitative differences between the various regions in more detail.

Figure 1.4: Long-Run Outcomes vs. Intertemporal Complementarity.

long-run bilingualism *long-run monolingualism*

$$\longleftarrow \text{REGION 1} \underset{0}{+} \text{REGION 2} \underset{\frac{\delta(\rho+\delta)(u_{\ell\ell}+u_{mm})}{\rho+2\delta}}{+} \text{REGION 3} \longrightarrow \left[u_{\ell L}+u_{mM}+\tfrac{u_{LL}+u_{MM}}{\rho+2\delta}\right]$$

adjacent complementarity *distant bilingualism*

Note: The figure shows how qualitative dynamics are categorized into three regions, based on the value of the formula in Definition 2.

1.3.2 Examples of Three Different Systems

To illustrate the different regions of Figure 1.4, consider the following simple example. Assume for simplicity that the production functions for both languages take the same Cobb-Douglas functional form, as in equation 1.8.

$$u\left(\ell, L, m, M\right) = \alpha \ell^{\gamma} L^{\beta} + (1-\alpha) m^{\gamma} M^{\beta} \tag{1.8}$$

In this case all outcomes are driven solely by the concavity of the Cobb-Douglas function and parameters δ, ρ, and α rather than by differences in languages themselves. The parameter α can be interpreted as the social environment, which determines how returns to using each language may differ even when the languages' production functions are identical.

For the purposes of illustration, this Cobb-Douglas function is particularly intuitive. It yields a unique interior steady state which is also a global maximum or saddle point in the feasible state space (see Result 2). This need not be true for all functional forms, and counterexamples are shown in Appendix B. In addition, long-run monolingualism and bilingualism nicely correspond to increasing and decreasing returns to scale. However, the general intuition developed from this example does not depend on the particular functional form.

Result 2 (Steady state properties from equation 1.8)**.**

With the Cobb-Douglas utility function in equation 1.8 the following are true:

1. *There is a unique interior steady state given by* $\ell^* = \frac{x}{1+x}$ *and* $L^* = \frac{\ell^*}{\delta}$*, where* $x = \left(\frac{1-\alpha}{\alpha}\right)^{\frac{1}{\gamma+\beta-1}}$.

2. *The steady state is a critical point of the constrained utility function in* (ℓ, L)*-space.*

3. *The steady state is a global maximum in the feasible state space if and only if the Cobb Douglas function shows decreasing returns to scale:* $\gamma + \beta \leq 1$*. Otherwise it is a saddle point.*

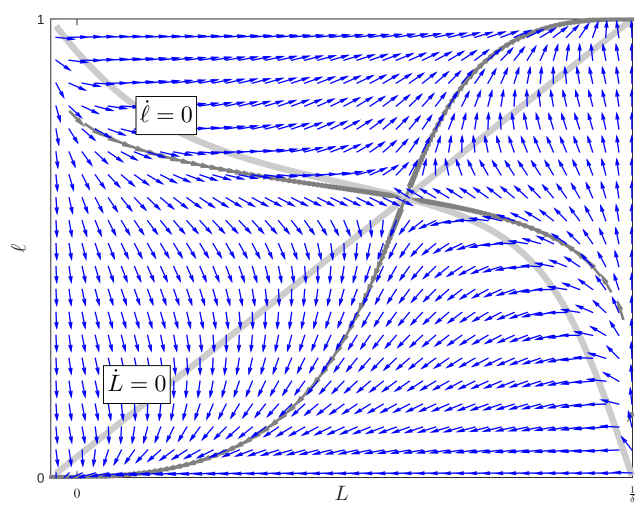
4. *The steady state is stable whenever* $\gamma + \beta < 1$*.*

Point (*4*) of the result provides a subtle mathematical caveat: it implies that, for this particular choice of utility function, a steady state cannot be unstable despite being a local maximum. In general, such a scenario could occur, e.g. the Jacobian and the Hessian could have opposite signs. But the general intuition behind unstable steady states would still apply, with the difference that an unstable local maximum may appear locally stable. In particular, in a region around the steady state trajectories will diverge very slowly. Therefore, omitting this case costs little in generality, although Section 3 provides further details on such systems.

To illustrate possible trajectories, suppose otherwise identical immigrants arrive at different ages; all have the same initial stock $L(0)$. Given their stock, person i chooses initial investment $\ell_i(0)$ that puts them on a deterministic trajectory according to differential equations 1.7. Immigrants' various ages are captured by their remaining lifespans T_i, so that the trajectories of ℓ_i and L_i are completely determined by u, $L(0)$, and T_i.

We now build intuition behind the dynamics of Regions 1, 2, and 3 of Figure 1.4. In Region 1, language is not addictive and the system has one stable manifold, as shown in Figure 1.5.[8] In fact, the steady state is a local maximum, so everyone (regardless of initial fluency $L(0)$) wishes to move toward that point. The loci $\dot{\ell} = 0$ and $\dot{L} = 0$ are graphed in light gray, and the two invariant manifolds are graphed in dark gray.

Figure 1.5: Bilingualism with Non-Addictive Language.



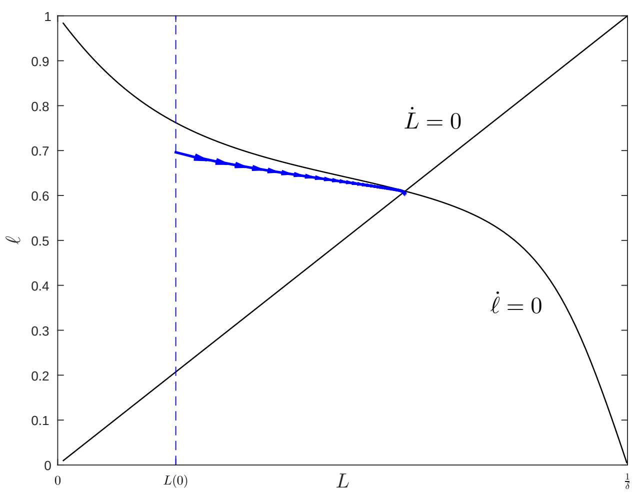
Note: The figure shows a system with a stable manifold and distant complementarity, corresponding to Region 3 in Figure 1.4. The fixed point is a utility maximum. Light gray lines indicate the loci $\dot{\ell} = 0$ and $\dot{L} = 0$. Dark gray lines show the stable and unstable manifolds. The parameters are $\alpha = 0.7$, $\gamma = 0.1$, $\beta = 0.1$, $\delta = 0.9$, and $\rho = 0.9$; they were chosen to maximize clarity of the graph.

Now consider the corresponding trajectories of ℓ and L. Every immigrant, regardless of the initial stock, will choose $\ell(0)$ to put themselves on the stable manifold, and then

8. The parameters are $\gamma = 0.1$, $\beta = 0.1$, $\alpha = 0.7$, $\delta = 0.9$, and $\rho = 0.9$, which were chosen for aesthetic reasons in illustrating the intuition, not for any empirical realism.

monotonically increase their utility as they move towards the steady-state. But the slope of the stable manifold implies that $\ell(t)$ and $L(t)$ will move in opposite directions, illustrating how $\mathcal{L}$ conforms to the definition of a non-addictive good. Figure 1.6 provides an illustration in which the individual begins with a moderately low level of stock $L(0)$. The solid blue line shows the trajectory; note that the individual invests more early in life, lowering $\ell(t)$ as fluency increases and eventually splitting time between languages more evenly than at first.

Figure 1.6: Bilingual Equilibrium Trajectory Without Addiction

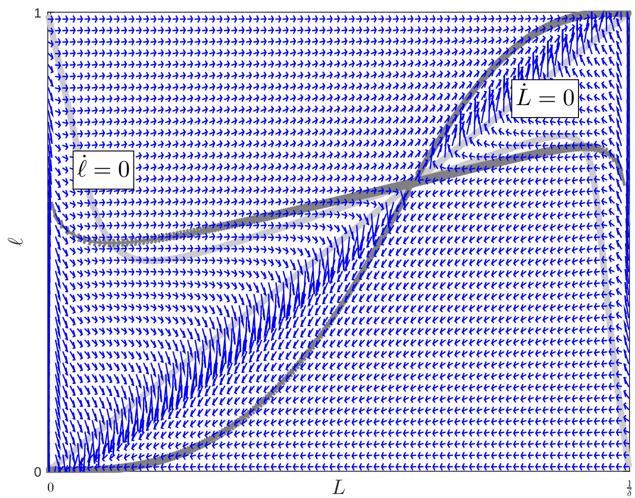


Note: The figure shows sample language trajectories for the system graphed in Figure 1.8. Regardless of the initial stock $L(0)$, individuals will choose $\ell(0)$ to put themselves on the stable manifold. They will then converge to the steady state. Note that $\dot{\ell} < 0$ while $\dot{L} > 0$, implying distant complementarity and non-addictive language.

In Region 2, the phase diagram subtly shifts, so that the process is addictive. Figure 1.7 shows such a system.[9] As in Region 1, the steady state yields the highest utility, so that people will always choose ℓ to be on the stable arm. But now the slope of the stable arm is positive, so that both $\ell(t)$ and $L(t)$ increase over time.

9. The parameters are $\gamma = 0.3$, $\beta = 0.3$, $\alpha = 0.9$, $\delta = 0.1$, and $\rho = 0.05$

Figure 1.7: Bilingualism with Addictive Language.

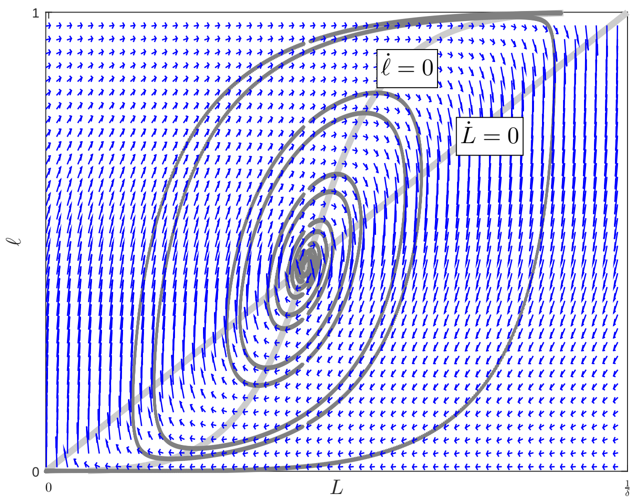


Note: The figure shows a system with a stable manifold and adjacent complementarity, corresponding to Region 2 in Figure 1.4. The fixed point is a utility maximum. Light gray lines indicate the loci $\dot{\ell} = 0$ and $\dot{L} = 0$. Dark gray lines show the stable and unstable manifolds. Note that on the stable arm near the steady state $\dot{\ell} > 0$ and $\dot{L} > 0$, implying adjacent complementarity and addictive language. The parameters are $\alpha = 0.9$, $\gamma = 0.3$, $\beta = 0.3$, $\delta = 0.101$, and $\rho = 0.05$; they were chosen to maximize clarity of the graph.

In the monolingual case of Region 3, the utility function changes concavity so that the steady state is a saddle point. This drastically changes the trajectories because the upper-right and lower-left corners of the state space yield the highest utility. Figure 1.8 shows the basic dynamics.[10] Both invariant manifolds are unstable, spiraling away from the steady-state towards the corners. Trajectories spiral around the steady state as well. In this case, people will eventually be completely monolingual in either $\mathcal{L}$ or $\mathcal{M}$ because the corners yield higher utility than any combination of $\mathcal{L}$ and $\mathcal{M}$ together. It is important to note that in this particular case, being fluent in language $\mathcal{L}$ is strictly better on a period-by-period basis than being fluent in $\mathcal{M}$. Yet the costs of attaining fluency in $\mathcal{L}$ may be too great, in a present-value sense, to allow for some people to choose that path. In this sense, the model fits the standard human capital framework: shadow prices determine whether individuals trade off present utility for higher future utility.

10. Parameters here are $\gamma = 0.7$, $\beta = 0.8$, $\alpha = 1$, and $\delta = 0.9$.

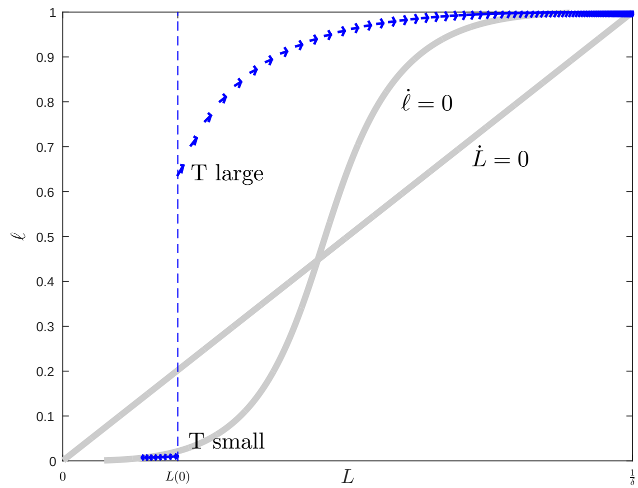
Figure 1.8: Monolingualism as a Steady State



Note: The figure shows a system with an unstable fixed point, which is a saddle point on the utility function. Hence all trajectories spiral towards the corners. Light gray lines indicate the loci $\dot{\ell} = 0$ and $\dot{L} = 0$. Dark gray lines show the two unstable manifolds. Full specialization in one language or the other is the result for any initial points other than the steady state. The parameters are $\alpha = 0.9$, $\gamma = 0.7$, $\beta = 0.8$, $\delta = 0.101$, and $\rho = 0.05$; they were chosen to maximize clarity of the graph.

To illustrate more explicitly the possible paths to monolingualism, consider Figure 1.9. Two people begin with the same stock $L(0)$. One person has long time horizon T_i, while the other has small T_i. The initial stock is low enough that speaking $\mathcal{L}$ generates less utility today than speaking $\mathcal{M}$, but the corner solution of $(\ell, L) = \left(1, \frac{1}{\delta}\right)$ generates higher utility than the other corner in the long run. Therefore, someone who is sufficiently young (or sufficiently patient) will find it worthwhile to invest in L, even though they are getting lower utility today than they would have otherwise. But an older person, or a sufficiently impatient one, will prefer to specialize in $\mathcal{M}$. The two blue lines show the respective paths.

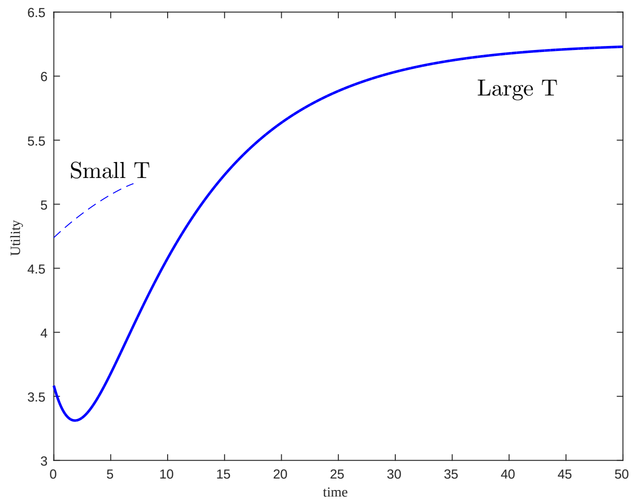
Figure 1.9: Monolingual Equilibrium Trajectories



Note: The figure shows sample language trajectories for the system graphed in Figure 1.5. Individuals choose to move to one of the corners, where utility is highest. Although utility is technically higher at the upper-right corner (full assimilation to local language), when initial stock $L(0)$ is small then individuals generate higher flow utility from using $\mathcal{M}$ than $\mathcal{L}$. Thus, only if the individual has a sufficiently long time horizon will it be worthwhile to invest in L, trading lower utility today for higher utility later (dotted line). When individuals have relatively short time horizons they specialize in $\mathcal{M}$ and converge to the other corner.

The intertemporal tradeoffs are illustrated in Figure 1.10, which graphs the utility period-by-period over the life span of each individual. Clearly, the older person chooses a path that is smoother but lower on average. The younger person actually has decreasing utility at first, until L becomes high enough that he can get much higher utility for the rest of his life.

Figure 1.10: Utility Over Time, Monolingual Case.



Note: The figure shows utility levels corresponding to paths shown in Figure 1.6.

Figure 1.9 shows that the policy function $\hat{\ell}\,(L,T)$ may not be continuous in either argument. Fixing $L(0)$, it is clear from the figure that different values of T may yield different trajectories. When some values of T imply $\hat{\ell}(L(0),T) > \delta L(0)$ and others imply $\hat{\ell}(L(0),T) < \delta L(0)$, there must be a threshold value $\hat{T}(L)$ for which the policy function is discontinuous. Similarly, fixing T, we may find that certain values of $L(0)$ imply $\hat{\ell}(L(0),T) > \delta L(0)$ while others imply $\hat{\ell}(L(0),T) < \delta L(0)$; again there will be a threshold $\hat{L}(T)$ for which the policy function is discontinuous.

There are important implications for behavior near these discontinuity thresholds. Small changes to a parameter – particularly α in this example, since α determines the location of the steady state in the feasible space – can cause someone to jump to the other side of the threshold, and their trajectory to switch direction. On the margin, then, trajectories could be very sensitive to events that affect α such as an influx of new speakers of $\mathcal{M}$. Trajectories could also be sensitive to increases in L, such as by virtue of taking a language course.

In addition, the illustration shows that in Region 3 people with the same stock $L(0)$ can opt for different levels of investment. This is not true in Regions 1 or 2, in which everyone follows the stable arm, regardless of other parameters. To allow for different investment behavior from people with similar fluency requires a mechanism like Region 3, in which the assimilation process is addictive.

1.4 Qualitative Evidence

The foregoing model applies neatly to language use, but can illuminate a variety of other cultural phenomena. Before turning to quantitative evidence for particular model predictions, I discuss some qualitative and anecdotal evidence from other fields of research.

1.4.1 Black Markets for Culture

Throughout history, people have practiced certain forms of cultural expression even when they are explicitly banned by the government. More interestingly, cultural practices that were not practiced are often revived once government policy changes. In terms of this model, such policies are basically sudden price changes, manifested as changes in α. Their effects should depend on whether they are expected or unexpected; temporary or permanent; and large or small. The model easily predicts both behaviors, modeling the beginning and end of government repression as sudden changes in the value of α. If the change is great enough, people will switch to the officially legal cultural practice; but if depreciation is slow enough and the ban is short-lived, there could be enough productive capital remaining that people will choose to revive the first practice once the ban ends.

Examples include Franco's suppression of the Catalan language in Spain, and Soviet suppression of religion. Catalan was explicitly outlawed, and although people continued to speak it at home it was not taught in schools or used in business or in the media (Boada, 2015; Mir, 2008). More compellingly, Soviet policies sought to suppress religion, often to great effect. For instance, religious rites dropped precipitously in certain regions, while officially-sanctioned atheistic rites rose (Froese, 2004, Table 3). Yet even after 70 years of suppression, Christianity sprang back to life after the end of Communism. To paraphrase Froese (2004), Spanish and Soviet policies had every competitive advantage, and yet they still failed to permanently convert people. If people only cared about having any language or any religion, we might expect that Catalan and Christianity would have been fairly quickly abandoned. That they were not indicates there is something durable about modes of communication and belief. The model makes this process explicit. If governments expect to change cultural norms or institutions, it may be more effective to do so by finding new uses for existing stocks of capital, rather than banning the stocks altogether.

1.4.2 Seemingly Maladaptive Norms

Some debate has surrounded the question of why certain cultural norms persist even when they have adverse economic consequences. In certain cases, ignorance to economic disadvantages hardly seems a legitimate explanation for the continuation of the norms in question. The model above can rationalize the behavior of people who adopt or maintain a cultural norm, despite having clear evidence of its economic inferiority: it may not be worth the short-run costs to develop a different cultural stock.

The Amish and other Anabaptist communities offer a modern example. It simply does not seem reasonable to claim that members of these sects are oblivious to the opportunities they forgo by staying in their community. This is especially true given that teenagers are allowed relatively unfettered access to the mainstream American community during *rumspringa*. Even still, 90% of the youth rejoin their childhood community (Schachtman, 2006). The model in this paper would rationalize Amish or Anabaptist communities as nurturing culture-specific capital that is difficult to transfer to mainstream America; it has a low α in mainstream society and a high α in the enclave. If M represents Amish cultural capital, it must be so unproductive in mainstream society that a young person on *rumspringa* does not find it worthwhile to bide his time depending on it while he develops a stock of mainstream cultural capital. Figure 1.9 encapsulates this case, where the average $L(0)$ is so low that only 10% of individuals choose a trajectory towards the upper-right corner. As Richerson

& Boyd (2005) point out, the strict separation from mainstream society, along with a special emphasis on religious rules and norms, is partly what ensures this low $L(0)$ and low α, thereby also ensuring the successful continuation of Anabaptist culture.[11]

A related phenomenon is when behaviors persist in a new environment, even when the behaviors seem sub-optimal in the new context. Atkin (2016) finds that Indian migrants prefer the food of their native region, giving up 7% of their caloric intake versus what they could alternatively buy in their destination region. Children's names in the African American community is another well-studied outcome; the negative labor market implications of black-sounding names are documented by Bertrand & Mullainathan (2004), and an identity explanation is proffered by Fryer & Levitt (2004). Such patterns could imply that tastes may reflect underlying identity stocks, for which non-pecuniary benefits explain the seemingly sub-optimal behavior.

1.4.3 Identity and Language Change

Intercultural adoption offers evidence that ethnic self-identity adapts to one's social context, but is also quite durable. Richerson & Boyd (2005, pp. 39-42) summarize the scant but suggestive work in this area. Notably, the results imply that individuals identify with their adoptive culture, and that the younger the adoptee, the less their interest in their native culture (Lydens, 1988; Andujo, 1988). Other evidence on the durability of cultural identity capital comes from Native American adoption of European settlers. Historical records document adoptees – taken captive by certain tribes between the ages of 5 and 12 – being re-integrated into their birth families decades later. The younger the child at the time of adoption, the more they identified with the adoptive tribe. Even after re-learning English, many tended to maintain their adoptive identity (Heard, 1973). There are several other factors at play, particularly the amount of racial discrimination in the adoptive community,

11. From Richerson's and Boyd's viewpoints, however, mainstream society is actually the maladaptive community, because they find no evidence that low birth rates of the developed world have a have a fitness-enhancing effect.

but Richerson and Boyd interpret the literature overall as suggesting that adoptees under the age of 10 tend to fully and permanently assimilate into the adoptive culture. While these patterns in young children may not be surprising, they offer evidence that identity can behave as a durable stock.

Language attrition also shows that culturally-specific characteristics can wax and wane over time. Schmid (2011, pp. 69-95) discusses the relevant literature, noting that several open questions remain. It may be true that native language fluency becomes much more permanent after puberty (Spångberg, 2009). But even in ethnic enclaves, language change, albeit not outright attrition, is very common. And regardless of the particular languages being considered, the reconfiguration of the enclave's linguistic syntax follows some standard patterns. Maher (1991), for example, documents such patterns in three immigrant and three indigenous enclaves.

The literature also relates language attrition to international adoption. For example, Koreans adopted by French families before age 10 had completely lost the ability to speak or even recognize Korean as adults (Ventureyra & Pallier, 2004). There may also exist a link between language attrition and identity formation. The wish to identify completely with a new community – particularly from others' point of view – can lead people to give up their native language rather quickly, trying to speak the second language with no accent. This might seem especially common with adolescent immigrants and child siblings (Schmid, 2011, p. 76), but studies linking language attrition to identity assimilation have not found a correlation with age at emigration (Schmid, 2002). If nothing else, the patterns offer evidence that one margin of assimilation can enable another in a causal way, which is an important topic for future study.

1.5 Conclusion

Current theories of social interactions generally assume individuals want to be like each other, and the models are set up so that investments made by one individual necessarily induce a corresponding investment in other individuals. The model presented in this chapter was based on a less restrictive premise: individuals don't care about being similar per se; rather, they care about having traits that allow them to engage in productive communication. Outside the context of cultural assimilation, this framework can speak to a wide variety of phenomena, especially the conditions under which individuals will endogenously form groups, and whether existing groups will endogenously become more homogeneous or more heterogeneous. There are numerous potential applications of a full-fledged human capital theory of group formation: understanding political polarization, modeling economic conditions behind cooperation and public good provision, or predicting changes in social norms and group identities. Incorporating social interactions into this model would be a major step forward in achieving such a theory.

While perhaps nontrivial to model such interactions, the individualistic theory presented here offers insight into the types of dynamics an interaction theory would need to model. First, it would need to allow for an individual's own capital stock to be either a substitute or complement for another individual's capital. Second, the intertemporal elasticities within and across individuals would need to be allowed to be positive or negative. The theory should be able to predict when productive interactions will be a function of individuals being different rather than being the same. Only then will be able to understand the true value of group membership, as opposed to individualistic identity, and to begin to understand the ways in which different aspects of individual identity are complementarity or not, how different personality traits are deployed in various circumstances for productive purposes, and how individuals invest in different personality or identity traits. In other words, the ultimate goal may be a theory of the Economy of Identity.

CHAPTER 2
EVALUATING THE DISTRIBUTION OF LONG-RUN ASSIMILATION OUTCOMES

2.1 Introduction

The first chapter outlined a model of cultural assimilation using the basic tools of human capital theory. As an empirical contribution, in this chapter I bring the theory to bear on several indicators of assimilation. The theory distinguished conditions that predict increasing homogeneity from those predicting increasing heterogeneity. In the empirics, I focus on testing for evidence of the latter. This makes two major points. First, any evidence in favor of the increasing-heterogeneity state of the world is, in turn, evidence that previous theories of assimilation were insufficient to model the process. Second, such evidence also points out the importance of higher moments in modeling the distribution of assimilation outcomes: increasing variance, as well as changing mean, characterize the increasing-heterogeneity system.

Previous studies have generally focused on group means, which are certainly important but only tell part of the story. When people assimilate at different rates (thereby producing a longer and longer tail in the distribution over time), it is those who assimilate slowly who are perhaps of greatest scientific and policy interest. Why are they assimilating slowly? Could any policy interventions improve individual or societal welfare by shifting incentivizing different assimilation rates?

Chapter 1 abstracted away from any single margin of assimilation, instead conflating many indicators as being driven by one underlying process. Thus the cultural stock can represent language fluency, religiosity, self-identity, or moral beliefs. I demonstrate the robustness of the model to these various measures of assimilation using three sociological surveys from different destination countries. The surveys are: the New Immigrant Survey (NIS) in the United States, the National Immigrant Survey (ENI) in Spain, and the Trajectories and Origins Survey (TeO) in France.[1] Each survey offers a great amount of detail on immigrants' migration trajectories and a variety of social outcomes – language skills, self-identity, religiosity, sense of belonging, and personal beliefs – while also recording characteristics at the time of migration – education, fluency, employment, wages, and marital status. When possible, these datasets are supplemented by larger-sample synthetic cohort analyses from the U.S. Census. Using the surveys, I test for various outcomes that distinguish the addictive- and non-addictive-culture states of the world, including: increasing variance in cultural outcomes, sorting by age and time since arrival, and bimodal versus unimodal long-run patterns.

Section 2.2 describes the data sets, while Section 2.3 walks through the evidence for different predictions of the model. Section 2.4 discusses potential future work.

2.2 Data

Different parameters of the model can generate different long-run distributions of human capital. It is an empirical question to determine which scenario (addictive/non-addictive and bilingual/monolingual) most closely explains observed outcomes. After establishing a candidate for the true assimilation mechanism, we can test predictions which follow uniquely from the case under consideration. I examine empirical evidence from four data sources. Three of the sources are sociological surveys containing information on a person's entire migration history: the New Immigrant Survey (NIS) from the United States, the National Immigrant Survey (ENI) from Spain, and the Trajectories and Origins survey (TeO) from France. These surveys record pre-immigration investments and information about initial stocks $L(0)$, as well as pre-migration work history, education, and marriage. Additionally, they provide descriptions of individuals' family structure, assets, loans to and from other family members, refugee status, and remittances. I focus particularly on the TeO survey because it provides the richest set of variables. In addition to those already mentioned, the TeO survey asks

1. To replicate the variables used in this paper, see the Data Dictionary available on the author's website.

about individuals' identity, sense of belonging, fluency in their native language, and experiences with discrimination. In addition, it provides information about neighborhood-level demographic characteristics: numbers of immigrants, unemployment rates, and average incomes. The final source is the U.S. Census, which is used for a few supplementary purposes. First, it provides the best historical evidence that assimilation patterns are largely similar across the last century. Second, it provides a baseline comparison to previous studies of immigration – particularly Borjas (2015), who uses a similar empirical analysis to examine average trajectories. Finally, it offers a large-sample check for the evidence from the other surveys.

The data need re-coding in order to ensure commensurability, and the online data dictionary (available on the author's website) has complete information for replicating all the variables used in this paper. However, some additional points should be made clear. First, to proxy for cultural distance, each survey was matched to linguistic proximity data. This requires knowledge of an individual's mother tongue, in order to calculate a distance to the relevant local language (English, Spanish, or French). The Census and ENI surveys explicitly collect data on mother tongues, but for the NIS and TeO surveys the mother tongue is imputed based on a combination of other information. After identifying each individual's mother tongue, linguistic proximity can be calculated using the method of Fearon (2003) and data from Ethnologue (Lewis *et al.* , 2016). Further details and formulas are provided in Appendix 1. To read the tables that follow, note that language proximity is normalized between 0 and 1, with 1 meaning that the languages share all their branches on the language tree and 0 indicating that they don't share any. A convexity parameter shapes the distance function so that languages with fewer branches are relatively farther apart than languages that share more.

The nature of the data presents a few serious problems. Surveys of language fluency and identity rely on self-reported answers, usually ranked on a Likert-type scale. Inference may be problematic when self-assessments are unreliable or incomparable across individuals.

Misclassification can lead to biased parameter estimates, as noted in Dustmann & van Soest (2001, 2004). Further, different individuals may differentially interpret the possible answer categories. For instance, people from different countries may have a different idea of what it means to speak English "very well." Further, different surveys sometimes have different scales. King *et al.* (2004) discuss the issue of comparability and suggest using vignettes as calibrating questions. However, that is not a possibility in my data.

Moreover, when the distribution itself is of interest (as opposed to the mean or a binary summary of the data), Likert scales are problematic even in the absence of misclassification error. Changes in the distribution can be mis-identified or misinterpreted for several reasons. In Chapter 3 I seek to mitigate several of these problems by constructing indices of language proficiency based on multiple types of survey questions. This combination of objective and subjective data provides not only a more-continuous measurement of fluency, but also lessens the bias due to misclassification or cross-cultural incommensurability. For now, I rely on the self-reported answers. This is the standard approach in essentially all studies of assimilation, and fortunately the nature of migration raises the likelihood of reliable answers. As Belli *et al.* (2001) emphasize, focusing on major life events minimizes measurement error. Hence, respondents are more likely to remember how well they spoke a language when they first arrived in a country than at another arbitrary calendar date.

Table 3.2 provides summary statistics of the variables used in each sample. There are some major differences in the populations sampled by different surveys, which are described in more detail below. Yet those differences serve as a check for the robustness of the model, by allowing for comparison between a broad range of different groups. Complete details for constructing and re-coding variables may be found in the online data dictionary.

Table 2.1: Summary Statistics by Survey

	U.S. Census 1980-2010 (N=3,072,227)	NIS 2003-2004 (N=7801)	ENI 2007 (N=15423)	TeO 2008-2009 (N=8734)
Age at Migration				
Mean	25	33	27	20
Median	23	30	26	21
Max	93	94	91	60
Tenure in Country				
Mean	16	6	12	21
Median	11	3	6	21
Max	90	64	89	59
0-5 years	24%	61%	47%	12%
6-10 years	18%	18%	20%	16%
11-15 years	14%	12%	7%	9%
16-20 years	11%	6%	5%	13%
21-30 years	14%	2%	8%	23%
31+ years	19%	1%	13%	27%
Demographics				
European	23%	16%	40%	28%
N. African/Mid. Eastern	2%	5%	13%	41%
Sub-Saharan	1%	9%	4%	15%
Latin American	44%	37%	40%	2%
South/East Asian	22%	32%	3%	13%
Female	51%	52%	55%	53%
Refugees	-	7%	-	4%
Local Language Skills				
Native Speaker	5%	6%	48%	13%
Speaks Well	73%	50%	95%	66%
Ever Spoke as Child	-	9%	-	29%
Spoke Well on Arrival	-	-	-	23%
Speaks Regional Lang. Well	-	-	20%	-
Education & Employment				
Less than H.S.	44%	34%	39%	49%
H.S./Some College	39%	20%	45%	34%
College Degree or More	17%	47%	21%	17%
Employed	51%	59%	64%	66%
In School	19%	2%	8%	3%

Note: Summary statistics are for immigrants only. Variables are described in Section 2.2. See online data dictionary for complete details on constructing variables from raw survey data.

2.2.1 U.S. Census

Census data comes from IPUMS, and consists of the 5% sample for 1980-2000. (Other available Census years lack information on English fluency, or have only binary measures which are not useful for the regressions I employ.) The sample is limited to immigrants, defined as anyone born abroad (not in Puerto Rico or on a Native American reservation, even though those groups could be culturally different from most Americans in ways similar to immigrants). This yields 3,072,227 individuals in 1980 or later. Unlike some previous studies, I do not limit the sample to working-age people, or to those who immigrated after a certain age. Since we wish to characterize the assimilation path over the whole lifecycle, it is important to include children to observe a "cutoff" date after which adaptation to the language and norms of a new country is no longer immediate or innate.

The first column of Table 3.2 summarizes the data, which essentially mirror the typical patterns of the so-called fourth wave of immigration (post-1965). These immigrants have generally arrived in their mid-twenties from Asia and Latin America more than Europe; only 5% come from English-speaking countries. They are about half female and about 20% report being in school at the time of the survey.

English ability takes one of five values, from "Does not speak English" to "Speaks only English." Again, unlike most past studies, I keep these categories rather than recoding fluency as a binary variable. I group individuals into year-of-arrival cohorts using the same definitions as in Borjas (2014, Chapter 2), and eight age-of-arrival bins: under 15, 16-20, 21-25, 26-30, 31-40, 41-50, 51-60, and over 60. For the synthetic cohort analyses that follow, I track age/year/birthplace cohorts using these groupings.

2.2.2 New Immigrant Survey: USA (NIS)

The New Immigrant Survey (NIS) is a nationally-representative survey of adults who obtained permanent legal resident (LPR) status in the United States (see Jasso *et al.* , 2005, for an overview of the survey). The respondents were surveyed in 2003 and 2004, and 99.5% had obtained permanent residency in 2003. As with the Census, I keep immigrants of all ages. I exclude immigrants not currently living in the United States (about 1.2%), and those for whom a full migration path cannot be verified (see the supplementary appendix for methodology behind defining the migration path). This leaves 7801 individuals.

By construction, the NIS is not representative of all immigrants. The selection is reflected in Table 3.2: immigrants in the sample arrive later in life, and have had shorter tenures, than immigrants from any other sample. Compared to the U.S. Census after 1980, NIS respondents are more likely to be Asian, less likely to be in school, more likely to work, and more likely to have a college degree. These differences reflect the path to LPR status: the main visas are for work authorization (hence the higher levels of employment and education in the NIS) or family reunification (hence the shorter tenures). NIS respondents may be older than Census respondents because they are more likely to have finished school (and now have a work visa) or to have adult children (who are U.S. citizens and sponsored the visa). In addition, refugees account for 7% of the NIS sample, while by definition there are no undocumented immigrants. In contrast, the Census does not allow for separate identification of refugees, yet contains an unknown proportion of undocumented immigrants.

English fluency is recorded on a 1-to-4 scale. The table shows that half of the respondents speak "well" or "very well." This is a lower proportion than any other survey, and may reflect the shorter tenure of the average respondent.

2.2.3 National Immigrant Survey: Spain (ENI)

The National Immigrant Survey (*Encuesta Nacional de Inmigrantes* or ENI) is a representative study of immigrants age 16 or older who have stayed or plan to stay in Spain for at least a year. Legal residence is not a consideration. People born abroad with Spanish citizenship since birth are excluded from the study, an important distinction for this study since such people may have many of the social and linguistic characteristics of non-citizen immigrants. The survey was conducted in 2007 with a sample size of 15,423 (details in Reher & Requena, 2009).

Immigrants to Spain are similar to those in the U.S. and France along several margins: they generally arrive in their mid-twenties, many are from Europe and Latin America, and their education levels are broadly similar. But compared to those in the USA and France, immigrants in Spain are more likely to come from North Africa, and much less likely to be Asian. This reflects both geographical proximity and Spain's relatively unpoliced oceanic borders near Africa. In addition, immigrants in Spain have shorter tenures than other countries, reflecting its different history of immigration: Spain experienced its first immigration boom only in the early 2000s. As of 2014, 12.8% of Spain's population was foreign-born, two-thirds of which were from outside the EU (Eurostat, 2015). In addition, immigrants from the former Spanish empire can obtain Spanish nationality after only two years, and in 2005 Spain offered amnesty to undocumented immigrants. Both these policies likely spurred increased migration from Latin America and northern Africa in the early twenty-first century.

Latin Americans account for two-fifths of the Census, NIS, and ENI surveys; yet a major difference between the samples, of course, is that most Latin American immigrants share a native language with Spain but not with the United States. Omitting all native Spanish-speakers drops nearly half the ENI sample. However, Spain also offers the unique opportunity to examine fluency in regional languages, where native Spanish speakers can be compared to others. There are three languages that are legally recognized as co-official languages in well-defined autonomous regions: Galician (Galego) in Galicia, Basque (Euskara) in the Basque Country, and Catalan (Catalá) in Catalonia, Valencia, and the Balearic Islands. Among the immigrants who live in one of these regions, 20% report speaking the local language "well" or better.[2]

The ENI records language fluency on a 1-to-5 scale. 95% of respondents speak "well" or better, the most of any survey; this is not just a product of the large proportion of native speakers (nearly half of the survey), since 87% of non-native speakers also report speaking Spanish well.

2. Interestingly, only 50% of these people are native Spanish speakers, indicating that the linguistic proximity of Spanish to Galician and Catalan may not be a factor in an individual learning the regional language.

2.2.4 Trajectories and Origins Survey: France (TeO)

The Trajectories and Origins Survey (*Trajectoires et Origins*, or TeO) is a survey of 21,761 individuals. The survey took place in 2008 and 2009 with a stratified sampling strategy to capture a representative sample from five different segments of the French population: immigrants (N=8734), children of immigrants (8241), migrants from French overseas departments (*départements d'outre-mer* or DOMs, N=725), children of DOM-ers (653), and children of natives (3408). Unless noted otherwise, I consider immigrants and DOM-ers. For more information on the TeO survey, see the project description (Beauchemin & Simon, n.d.).

Like the USA and Spain, immigrants to France show some idiosyncratic patterns. They arrive at younger ages than immigrants in other surveys (early twenties); they have quite long tenures (21 years on average); and they are much more likely to be African or Middle Eastern, partly reflecting the French colonial history of many of these source countries. Immigrants to France are less educated than those in the US or Spain, with half having less than a high school degree and less than 20% having a college degree. Yet they are more likely to be employed, with fully two-thirds reporting having a job at the time of the survey.

The TeO survey provides the most detail on pre-migration characteristics. Nearly one third reported using French as a child, with 13% reporting it as their mother tongue. Fluency is coded on a 1-to-4 scale; 66% report speaking "well" or "very well" at the time of the survey, but only one-quarter reported speaking well on arrival.

2.3 Empirical Evidence for Long-Run Cultural Specialization

The most straightforward empirical exercise is to determine which mechanism best describes the data (multiculturalism or specialization). The correlation between language fluency and use (Figure 1.3) already presented some evidence for the addictive/specialization case. Here we proceed to proceed to examine further predictions that are specific to this case.

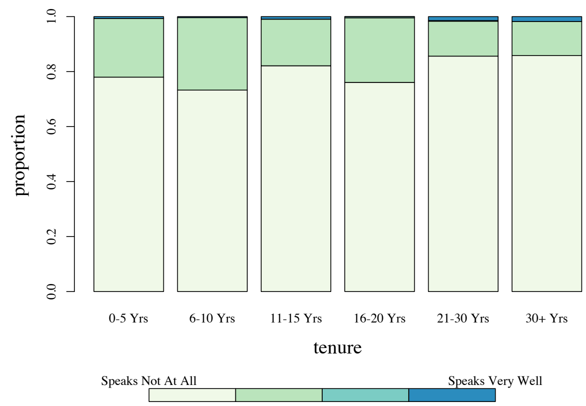
2.3.1 Prediction 1: Assimilation Trajectories Should Diverge

When the assimilation mechanism yields long-run monolingualism, we should observe individuals' trajectories diverging over time. Figure 1.9 showed this divergence for individuals of different ages; in general, the trajectories may diverge for several reasons, including different values of α. We should not be surprised to find diverging assimilation paths even among individuals of the same age at arrival. Using language fluency as the primary outcome of interest, I next examine evidence for this prediction.

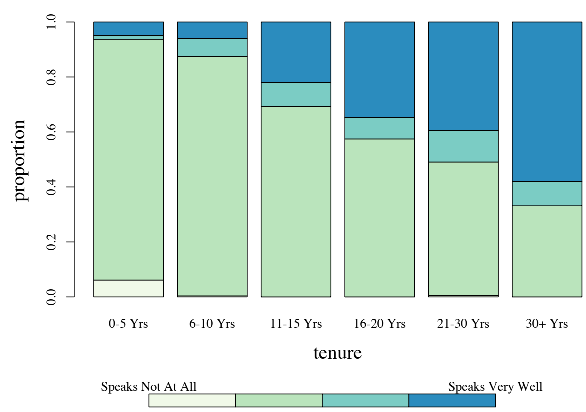
The French data provide a simple but interesting argument that cultural assimilation reflects an increasing-heterogeneity (as opposed to increasing-homogeneity) process. This is the only survey to detail French fluency at arrival as well as at the time of the survey. Panel (a) of Figure 2.1 shows the fluency of immigrants at arrival, based on time since migration. Although the answers are retrospective, and although the arrival cohorts may be quite different demographically, the distribution of fluency is quite similar. Almost no one reported speaking French well, and the vast majority knew none at all. Panel (b) shows how those fluencies changed by the time of the survey. In particular, most people learn some French within a few years, but many people (upwards of 25%) never learn more than that. This supports the position that linguistic assimilation is of the addictive type, at least at the aggregate, population-wide level. In fact, the data reflect Figure B.1, in which a stable steady-state induces some people to learn only a little bit of French, but most people are drawn to the upper-right corner and become mainly fluent (albeit reaching the corner at different rates).

Figure 2.1: Distribution of French Fluency for Turkish Immigrant Cohorts in TeO Survey.

(a) On Arrival



(b) At Time of Survey (2009)



Note: Figures show fluency in French, as reported on a 4-point scale, by cohort. Time in France is indicated on the x-axis; the panels capture changes in the distribution of reported fluency between arrival and the time of the survey.

However, the ideal test for the diverging-trajectory predictionwould be to track identical immigrants over time. Lacking this possibility in the data, I perform a synthetic cohort analysis on age-year-birthplace cohorts in the U.S. Census. Previous studies, particularly Borjas (2015), use birthplace/tenure/age at arrival cohorts. For my purposes, it is important to condition on the distribution of fluency at arrival. The 1980 Census roughly captures the initial distribution of language fluency for the 1975-1979 arrival cohort, and the 1990 Census roughly captures the same for the 1985-1989 arrival cohort. Comparing to the 2000 Census shows how age/birthplace groups have assimilated over one to two decades.

To perform this and subsequent analyses, I employ a location-scale ordered choice regression(Williams, 2009; Greene & Hensher, 2010). This test overcomes problems with simple logit and probit models, namely, bias in the presence of heteroskedasticity. The need to correct for conditional heteroskedasticity was noted by Allison (1999), and the location-scale model has become commonplace in a variety of applications. However, in this study the heteroskedasticity is itself an object of interest, since increasing within-group variance is an empirical prediction: namely, increasing variance is one manifestation of diverging trajectories.

In survey data, language fluency is reported on a Likert-scale with several ordered categories of outcomes. But the underlying dependent variable Y_i^* is continuous, and depends on observables X and Z according to equation 2.1.

$$Y_i^* = X_i'\beta + \sigma\left(Z_i'\phi\right)\epsilon_i \tag{2.1a}$$

$$\epsilon_i \sim N(0,1) \tag{2.1b}$$

Y_i^* is only observed discretely, with its observed counterpart Y_i taking the form of equation 2.2.

$$Y_i = 1 \text{ if } Y_i^* > \eta \tag{2.2}$$

The mean and variance are determined by X_i and Z_i, which may (but do not have to) share the same covariates. The conditional log-variance is directly proportional to $Z_i'\phi$:

$$\text{Var}\left[\epsilon_i \middle| Z_i\right] \propto \exp\left(Z_i'\phi\right)^2 \tag{2.3}$$

Therefore, the variance increases with variable z_{ij} whenever $\phi_j > 0$.

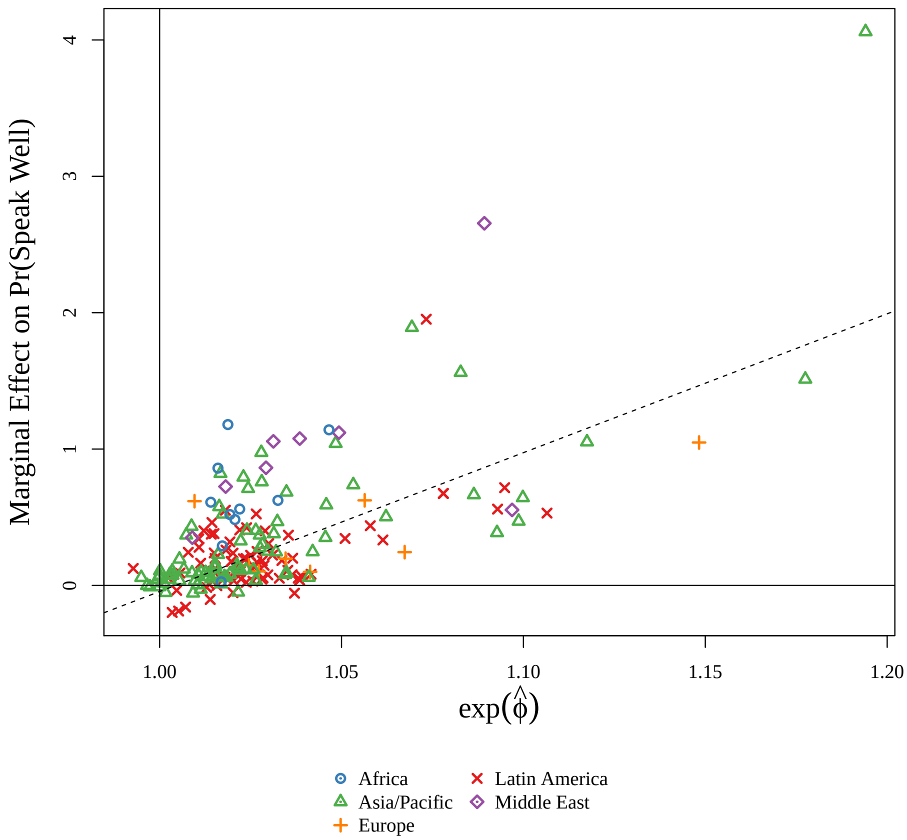
With the Census data, I run a location-scale probit on age/birthplace cohorts defined above, for those who immigrated between 1975 and 1979 or 1985-1989. Birthplace groups are given by the standard Census birthplace codes, dropping predominantly-English-speaking source countries.[3] Tenure is the single regressor in both the mean and variance equations.

Figure 2.2 plots the effect of an extra year of tenure along two dimensions. On the vertical axis is the change in the probability of speaking English very well – that is, being in the highest category of fluency. On the horizontal axis is the factor by which the variance grows or shrinks – that is, $\exp\left(\hat{\phi}\right)$. Cohorts are color-coded by region of birth. The dotted line shows a weighted regression line, with weights given by total group size across both Census surveys. Overall, regardless of birth region, Figure 2.2 shows a clear trend: groups that assimilate faster also have larger variance over time. This implies that, on average, most people assimilate, but a sizable left tail does not, or at least assimilates more slowly.

Figure 2.2 shows two other notable patterns. First, nearly all groups fall squarely in the upper-right quadrant, or at least near the axes. This affirms that, as a rule, groups tend to assimilate on average, but do so at different rates, leaving a lower tail that either does not assimilate at all or assimilates relatively slowly. This pattern is consistent with the long-run monolingualism scenario in the model, but not with any other scenarios. Second, many groups cluster around the origin, indicating that there is little change in their fluency distributions after immigration. This is not inconsistent with the model's predictions in the monolingual case; however, there are multiple reasons a group would not assimilate further after arrival. It could be that the group was nearly fully fluent at arrival, so that further improvement in English was not possible. It could also be that the group was older or tended to cluster in enclaves, and therefore had less incentive to learn English. Subsequent empirical exercises will attempt to disentangle age and arrival fluency.

3. The dropped birthplaces are: Canada, England, Scotland, Wales, UK, Ireland, Gibraltar, Australia, New Zealand, Bermuda, Belize, Jamaica, Anguilla, Antigua-Barbuda, Bahamas, Barbados, Virgin Islands, Cayman Islands, Dominica, Grenada, St. Kitts & Nevis, St. Lucia, St. Vincent, Trinidad & Tobago, Turks & Caicos, and Aruba.

Figure 2.2: Location-Scale Probit: Marginal Effects of an Extra Year in a Country, by Birthplace/Year of Arrival/Age at Arrival Cohort



Note: The figure shows the marginal effect of an extra of tenure: the x-axis is the multiplicative factor on the within-group variance; the y-axis is the marginal change in the probability of speaking English well. Dots represent a birthplace/age/year of arrival cohort and are color-coded by birthplace region. The dotted line is a weighted regression line using size of the group as regression weights.

The NIS, ENI, and TeO surveys also provide baseline evidence for the monolingual scenario. The surveys are too small to break into age- or year-of-arrival cells, so I use the same location-scale ordered probit model as above, applied to the surveys as a whole. Regressors consist of a quadratic polynomial in age at migration (AaM) and tenure, as well as an indicator for gender interacted with tenure. The results show multiple ways in which the mechanism underlying a J-shaped distribution might appear in the data.

Table 2.2 shows the baseline regression of language fluency on tenure and age at migration. Results for each survey are shown in a different panel, and each column controls for a different set of pre-migration characteristics: education before migration, birthplace, and linguistic proximity. Panel (a) shows the coefficients for regressions on NIS data. The coefficients predicting the mean align with model predictions, and are robust to the control variables: age at arrival is negatively correlated with fluency, tenure is positively correlated. Fluency is concave in both. The interaction age$\times$tenure is another measure of changing distribution of fluency; the negative sign implies increasing variance over time. Women have lower levels of English fluency, and accumulate fluency more slowly.

In the variance equation, positive coefficients imply increasing within-group variance; the table shows that fluency tends to increase in variance over time. As a whole, women are less heterogeneous than men, but their variance increases even more with time. These patterns are robust to including education at arrival, birthplace, and linguistic proximity to English.

The results are largely the same in Panel (b), which shows regressions on immigrants to Spain. The signs of all variables are the same as for U.S. immigrants, except women tend to be more fluent than men and also more hetergeneous.

Panel (c) shows the parameter estimates for immigrants to France. The coefficients for mean fluency demonstrate the same pattern as before, in terms of the signs and robustness to controls. The TeO survey has the richest set of variables, and records immigrants' French fluency on arrival. Thus, Column 5 of Panel (c) controls for the most characteristics of any regression: education on arrival, birthplace, linguistic proximity of mother tongue, and fluency in French. Even in this regression, the coefficients are robust. Females have lower levels of fluency than men (although not significantly so).

For the variance equation, however, the TeO data appear to have the opposite pattern from NIS and ENI: variance is decreasing over time (even more so for female, although the effect is negligible). This does not contradict the mechanism behind the J-shaped distribution. Rather, it implies that, overall, immigrants to France are fully assimilating. The mean equation says fluency is improving over time, while the variance equation says that people are not being left behind in the lower tail. The story, then, is consistent with the patterns in Figure 1.8 in the case where every trajectory leads toward the upper-right corner.

Table 2.2: Location-Scale Probit: Conditioning on Arrival Characteristics

(a) NIS

	(1)	(2)	(3)	(4)	(5)
Mean					
Age at Migration (AaM)	-0.0508***	-0.145***	-0.0863***	-0.163***	-0.167***
	(0.0122)	(0.0145)	(0.0132)	(0.0155)	(0.0164)
$(\text{AaM})^2/100$	-0.00422	0.131***	0.0358**	0.148***	0.156***
	(0.0129)	(0.0157)	(0.0138)	(0.0167)	(0.0177)
Tenure	0.102***	0.215***	0.195***	0.317***	0.329***
	(0.0292)	(0.0316)	(0.0308)	(0.0359)	(0.0384)
$\text{Tenure}^2/100$	-0.00444***	-0.00490***	-0.00548***	-0.00683***	-0.00667***
	(0.000968)	(0.000881)	(0.000935)	(0.000961)	(0.00105)
(AaM)×Tenure	-0.000803	-0.00269***	-0.00126*	-0.00337***	-0.00363***
	(0.000545)	(0.000674)	(0.000579)	(0.000737)	(0.000749)
Female	-0.345***	-0.165*	-0.278***	-0.0819	-0.182*
	(0.0757)	(0.0791)	(0.0785)	(0.0832)	(0.0919)
Female×Tenure	-0.00426	-0.0142	-0.00657	-0.0207*	-0.0144
	(0.00883)	(0.00933)	(0.00851)	(0.00958)	(0.00993)
Variance					
Tenure	0.00911*	0.00984*	0.00387	0.0105*	0.00890
	(0.00417)	(0.00395)	(0.00459)	(0.00426)	(0.00464)
Female	-0.0570	-0.0540	-0.0697	-0.0888*	-0.0599
	(0.0453)	(0.0445)	(0.0463)	(0.0452)	(0.0511)
Female×Tenure	0.0154**	0.0152**	0.0197***	0.0190***	0.0174**
	(0.00555)	(0.00529)	(0.00590)	(0.00551)	(0.00575)
Educ. on Arr.	No	Yes	No	Yes	Yes
Birthplace	No	No	Yes	Yes	Yes
Ling. Proxim.	No	No	No	No	Yes
N	7213	7001	7213	7001	5644

Note: Coefficients are from an ordered location-scale probit regression of a 4-category ordinal measure of English fluency: speaks ...
SOURCE: NIS survey, using all immigrants who are non-native English speakers. Linguistic proximity to English was calculated according to method of Fearon (2003).
* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$

(Table 2.2, continued)

(b) ENI

	(1)	(2)	(3)	(4)	(5)
Mean					
Age at Migration (AaM)	-0.0354***	-0.0495***	-0.0417***	-0.0551***	-0.0549***
	(0.00877)	(0.00908)	(0.00875)	(0.00897)	(0.0141)
$(\text{AaM})^2/100$	0.0209	0.0332**	0.0278*	0.0418***	0.0401*
	(0.0108)	(0.0111)	(0.0108)	(0.0110)	(0.0172)
Tenure	0.111***	0.110***	0.139***	0.133***	0.136***
	(0.0123)	(0.0126)	(0.0144)	(0.0142)	(0.0215)
$\text{Tenure}^2/100$	-0.120***	-0.121***	-0.162***	-0.156***	-0.130***
	(0.0136)	(0.0140)	(0.0161)	(0.0161)	(0.0249)
(AaM)×Tenure	-0.000610**	-0.000458*	-0.000794***	-0.000638**	-0.000757*
	(0.000220)	(0.000225)	(0.000231)	(0.000234)	(0.000341)
Female	0.125*	0.0715	0.0315	0.00176	0.0965
	(0.0582)	(0.0577)	(0.0584)	(0.0582)	(0.0969)
Female×Tenure	-0.00625	-0.00778	-0.00951	-0.00883	-0.0177
	(0.00570)	(0.00562)	(0.00595)	(0.00580)	(0.0103)
Variance					
Tenure	0.00637*	0.00724*	0.00834**	0.00856**	0.0155**
	(0.00320)	(0.00333)	(0.00320)	(0.00325)	(0.00474)
Female	0.117*	0.0962	0.0775	0.0715	0.154*
	(0.0507)	(0.0507)	(0.0512)	(0.0508)	(0.0745)
Female×Tenure	-0.000746	-0.00191	-0.00223	-0.00255	-0.0127*
	(0.00384)	(0.00392)	(0.00373)	(0.00375)	(0.00543)
Educ. on Arr.	No	Yes	No	Yes	Yes
Birthplace	No	No	Yes	Yes	Yes
Ling. Proxim.	No	No	No	No	Yes
N	6882	6882	6882	6882	3030

Note: Coefficients are from an ordered location-scale probit regression of a 5-category ordinal measure of Spanish fluency:
SOURCE: ENI survey, using all immigrants who are non-native Spanish speakers. Linguistic proximity to Spanish was calculated according to method of Fearon (2003).
* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$

(Table 2.2, continued)

(c) TeO (Full Sample)

	(1)	(2)	(3)	(4)	(5)
Mean					
Age at Migration (AaM)	-0.0536***	-0.123***	-0.0585***	-0.148***	-0.167***
	(0.0111)	(0.0118)	(0.0123)	(0.0138)	(0.0165)
$(\text{AaM})^2/100$	0.0608***	0.156***	0.0703***	0.203***	0.213***
	(0.0155)	(0.0170)	(0.0173)	(0.0200)	(0.0235)
Tenure	0.0359*	0.0359**	0.0487**	0.0576***	0.0923***
	(0.0150)	(0.0136)	(0.0161)	(0.0147)	(0.0174)
$\text{Tenure}^2/100$	-0.0607**	-0.0613***	-0.0724***	-0.0798***	-0.143***
	(0.0185)	(0.0171)	(0.0199)	(0.0184)	(0.0229)
(AaM)×Tenure	-0.000756*	-0.000196	-0.00105**	-0.000586	-0.00101*
	(0.000347)	(0.000299)	(0.000375)	(0.000329)	(0.000405)
Female	-0.126*	-0.0663	-0.117	-0.0304	-0.0282
	(0.0604)	(0.0607)	(0.0642)	(0.0628)	(0.0758)
Female×Tenure	0.00344	0.00319	0.00339	0.00241	0.00238
	(0.00253)	(0.00246)	(0.00267)	(0.00247)	(0.00277)
Variance					
Tenure	-0.0135***	-0.0148***	-0.00980**	-0.00931***	-0.00575*
	(0.00328)	(0.00261)	(0.00313)	(0.00249)	(0.00247)
Female	0.0631	0.0164	0.145	0.0175	0.0449
	(0.0703)	(0.0691)	(0.0744)	(0.0741)	(0.0725)
Female×Tenure	-0.00239	-0.00211	-0.00351	-0.00342	-0.00253
	(0.00339)	(0.00313)	(0.00330)	(0.00317)	(0.00294)
Education	No	Yes	No	Yes	Yes
Birthplace	No	No	Yes	Yes	Yes
Ling. Proxim.	No	No	No	No	Yes
Fluency at Migration	No	No	No	No	Yes
N	7601	7576	7601	7576	7282

Note: Coefficients are from an ordered location-scale probit regression of a 4-category ordinal measure of French fluency:
SOURCE: TeO survey, using all immigrants who are non-native French speakers. Linguistic proximity to French was calculated according to method of Fearon (2003).
* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$

2.3.2 Prediction 2: Assimilation is Monotonic by Age at Arrival

The baseline patterns generally imply that assimilation conforms to patterns modeled by Region 3 in the model, where long-run monolingualism is the rule. In this case, culture should be addictive, sufficiently so that trajectories diverge and a bifurcated distribution forms over time. We can test other predictions of the model, which should only hold under these particular conditions. Affirming these predictions provides a test of the particular mechanism outlined in this paper.

The first prediction is that the extent of a cohort's bimodality should be monotonic in the cohort's age at arrival. Older immigrants have less incentive to assimilate; hence, conditional on fluency at arrival, the lower tail of the long-run distribution should be larger when cohorts arrive at older ages. Conversely, the upper tail should be larger when cohorts arrive at younger ages.

To test for these specific changes in distributions I perform further synthetic cohort analysis on Census data, controlling for cohorts' fluency at arrival. I focus on 1975-1979 arrivals in the same age brackets as before, and compare the distribution of fluency in 1980 versus 2000. To control for the distribution of fluency at arrival, I calculate the proportional change in the CDF relative to the CDF in 1980. Specifically, suppose the possible fluency categories are ordered $n = 1, 2, \ldots, N$. Let the fraction of respondents in category n at time t be $f_t(n)$, and the CDF of the distributionbe $F_t(n)$. We would not observe inframarginal changes among those who were already in the top category $n = N$, so the relevant population is really the proportion $F_{1980}(N-1)$ of respondents who could potentially move to a higher category. Equation 2.4 gives the calculations for proportional changes in the CDF.

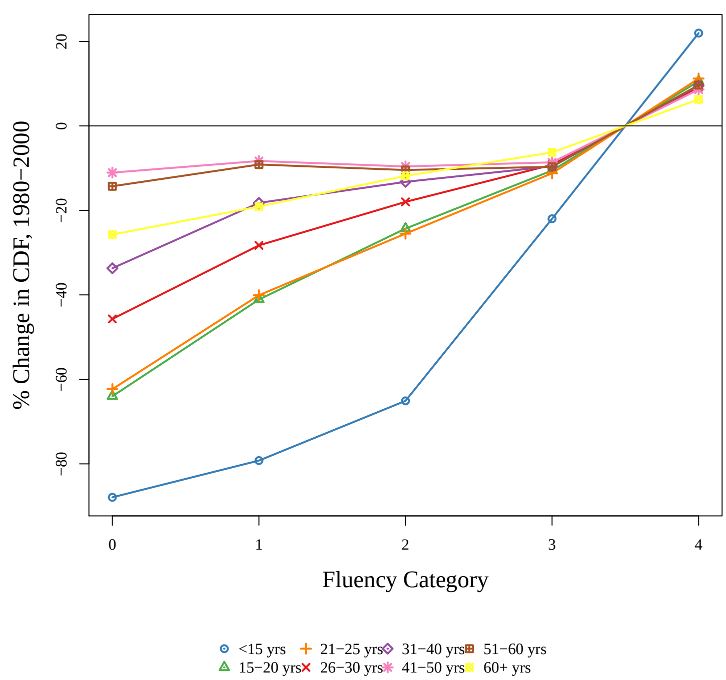
$$\Delta_n = \begin{cases} \frac{F_{2000}(n) - F_{1980}(n)}{F_{1980}(n)} & n \leq N-1 \\ \frac{f_{2000}(N) - f_{1980}(N)}{F_{1980}(N-1)} & n = N \end{cases} \tag{2.4}$$

We expect Δ_1 and Δ_N to be smaller in magnitude for older cohorts, since older immigrants have less incentive to learn the language. We also expect $\Delta_1 < 0$ for all ages, since at least someone will probably transition out of the lowest category; similarly $\Delta_N > 0$ for all ages, since at least someone will become very good at English. For the categories in between, the prediction is not so obvious: Δ_n could be positive or negative, depending on how quickly people are learning English, and whether they are assimilating all the way into maximal fluency or stopping at some mid-range level. But based on the hypothesis of a

J-shaped distribution, we would expect these to probably be negative, since everyone should be eventually grouping in the highest and lowest categories.

Figure 2.3 plots Δ_n for different age-of-arrival groups. Overall, the patterns confirm the sorting predicted by the model. For the lowest fluency category, the groups are ordered almost monotonically by age. The oldest immigrants (over age 50) are the only exception, having greater transitions rates out of this category than immigrants in their 40's. For the highest fluency category, the results are similar: immigrants in their 50's break the otherwise monotonic pattern in terms of transitioning into this category.

Figure 2.3: Change in Distribution of Fluency, by Age-Group



Note: The figure shows the change in the cumulative distribution of fluency for immigrants arrival between 1975 and 1979. Changes are calculated according to equation 2.4.

2.3.3 Prediction 3: Price "Experiments" Show Different Rates of Assimilation

Certain groups of immigrants should face different costs of assimilation, based on pre-migration investments in language skills. Immigrants who used a country's local language as a child (albeit not as a mother tongue) should speak the local language better and have different rates of assimilation than others; similarly for those who took a class before immigration. Refugees, on the other hand, likely did not make the same investments as others, and should have lower levels of fluency, with faster rates of assimilation.

Table 2.3 expands upon the baseline regressions by distinguishing these subgroups in the NIS and TeO surveys. Panel (a) shows the coefficients for the U.S. immigrants. The overall patterns for the mean are the same as before. Those with exposure to English do have better skills, but slower rates of acquisition. Refugees have worse skills, but higher rates. The evidence for heteroskedasticity is weak. The overall within-group variance appears to increase over time, conditional on the variables in the regression. Refugees, as a whole, appear to be less heterogeneous than average, but their within-group variance increases with time. Those who took an English class before immigration are also more homogeneous than average, but there is virtually no time effect on heteroskedasticity. Those who spoke English as a child do not have any group-specific variance effects.

Panel (b) of Table 2.3 shows the same coefficients for immigrants to France. The coefficients are exactly the same in sign, across the board, showing consistent patterns of linguistic assimilation for all groups, including refugees and those with previous exposure to French. Not only do the tables affirm the model separately for the USA and France, but they support the idea of a "universal" linguistic assimilation process, independent of an immigrant's host society.

Table 2.3: Location-Scale Probit: Conditioning on Pre-Migration Investments

(a) NIS

	(1)	(2)	(3)	(4)	(5)
Mean					
Age at Migration (AaM)	-0.0346**	-0.118***	-0.0644***	-0.133***	-0.138***
	(0.0126)	(0.0149)	(0.0136)	(0.0157)	(0.0163)
$(\text{AaM})^2/100$	-0.00224	0.110***	0.0287*	0.123***	0.132***
	(0.0132)	(0.0159)	(0.0139)	(0.0167)	(0.0175)
Tenure	0.156***	0.259***	0.244***	0.353***	0.373***
	(0.0336)	(0.0365)	(0.0362)	(0.0400)	(0.0400)
$\text{Tenure}^2/100$	-0.00494***	-0.00568***	-0.00619***	-0.00748***	-0.00745***
	(0.00110)	(0.00107)	(0.00111)	(0.00113)	(0.00114)
(AaM)×Tenure	-0.00164**	-0.00310***	-0.00200**	-0.00365***	-0.00394***
	(0.000578)	(0.000703)	(0.000617)	(0.000745)	(0.000736)
Female	-0.322***	-0.220**	-0.284***	-0.147	-0.226*
	(0.0734)	(0.0771)	(0.0771)	(0.0803)	(0.0886)
Female×Tenure	-0.00203	-0.00761	-0.00462	-0.0140	-0.00996
	(0.00857)	(0.00928)	(0.00857)	(0.00946)	(0.00973)
Refugee	-0.501***	-0.388*	-0.656***	-0.415**	-0.555**
	(0.132)	(0.154)	(0.141)	(0.147)	(0.200)
Refugee×Tenure	0.0849***	0.0739***	0.0594***	0.0402*	0.0689**
	(0.0181)	(0.0210)	(0.0164)	(0.0173)	(0.0254)
Used Eng. at Age 10	1.606***	1.653***	1.390***	1.401***	1.465***
	(0.163)	(0.176)	(0.156)	(0.168)	(0.205)
(Eng. Age 10)×Tenure	-0.0438*	-0.0375	-0.0360*	-0.0286	-0.0332
	(0.0216)	(0.0219)	(0.0181)	(0.0204)	(0.0234)
English Class	1.173***	0.935***	1.107***	0.874***	0.946***
	(0.0890)	(0.0879)	(0.0921)	(0.0901)	(0.0993)
(Eng. Class)×Tenure	-0.0124	-0.0391***	-0.0177	-0.0400***	-0.0506***
	(0.00952)	(0.0103)	(0.00947)	(0.0103)	(0.0101)
Variance					
Tenure	0.00903	0.0125**	0.00687	0.0134**	0.0124*
	(0.00524)	(0.00481)	(0.00528)	(0.00491)	(0.00518)
Female	-0.0793	-0.0588	-0.0928*	-0.0954*	-0.0636
	(0.0451)	(0.0459)	(0.0463)	(0.0466)	(0.0508)
Female×Tenure	0.0152**	0.0152**	0.0195**	0.0193**	0.0171**
	(0.00571)	(0.00563)	(0.00599)	(0.00587)	(0.00561)
Refugee	-0.467***	-0.344**	-0.359***	-0.397***	-0.377**
	(0.110)	(0.114)	(0.108)	(0.106)	(0.132)
Refugee×Tenure	0.0193	0.0144	0.0131	0.0108	0.0179
	(0.0153)	(0.0155)	(0.0148)	(0.0140)	(0.0171)
Spoke Any Eng. Age 10	-0.0458	0.0237	-0.0563	-0.00244	0.0181
	(0.0867)	(0.0862)	(0.0896)	(0.0892)	(0.0952)
(Eng. Age 10)×Tenure	0.00725	0.00375	0.00567	0.00474	0.00602
	(0.0122)	(0.0117)	(0.0126)	(0.0118)	(0.0122)
English Class	-0.217***	-0.120*	-0.155**	-0.111*	-0.0696
	(0.0480)	(0.0490)	(0.0501)	(0.0504)	(0.0525)
(Eng. Class)×Tenure	0.000925	-0.00336	-0.00240	-0.00454	-0.0120*
	(0.00604)	(0.00622)	(0.00664)	(0.00670)	(0.00592)
Educ. on Arr.	No	Yes	No	Yes	Yes
Birthplace	No	No	Yes	Yes	Yes
Ling. Proxim.	No	No	No	No	Yes
N	7040	6839	7040	6839	5644

Note: Regressions are similar to Table 2.2, Panel (a).
* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$

(Table 2.3, continued)

(b) TeO

	(1)	(2)	(3)	(4)	(5)	(6)
Mean						
Age at Migration (AaM)	-0.0448***	-0.112***	-0.0501***	-0.132***	-0.130***	-0.160***
	(0.0123)	(0.0126)	(0.0130)	(0.0138)	(0.0137)	(0.0165)
$(\text{AaM})^2/100$	0.0522**	0.144***	0.0594**	0.179***	0.178***	0.202***
	(0.0171)	(0.0177)	(0.0182)	(0.0197)	(0.0197)	(0.0233)
Tenure	0.0502**	0.0505***	0.0535**	0.0606***	0.0621***	0.0900***
	(0.0162)	(0.0144)	(0.0167)	(0.0148)	(0.0147)	(0.0172)
$\text{Tenure}^2/100$	-0.0743***	-0.0745***	-0.0762***	-0.0805***	-0.0838***	-0.137***
	(0.0201)	(0.0180)	(0.0210)	(0.0186)	(0.0194)	(0.0225)
(AaM)×Tenure	-0.00103**	-0.000446	-0.00113**	-0.000657*	-0.000713*	-0.00100*
	(0.000370)	(0.000314)	(0.000385)	(0.000333)	(0.000331)	(0.000401)
Female	-0.169**	-0.108	-0.158*	-0.0611	-0.0779	-0.0320
	(0.0631)	(0.0624)	(0.0641)	(0.0622)	(0.0622)	(0.0742)
Female×Tenure	0.00454	0.00413	0.00418	0.00304	0.00366	0.00234
	(0.00271)	(0.00257)	(0.00275)	(0.00247)	(0.00247)	(0.00271)
Refugee	-0.335**	-0.161	-0.354***	-0.190	-0.175	-0.0327
	(0.105)	(0.103)	(0.104)	(0.104)	(0.105)	(0.138)
Refugee×Tenure	0.00712	-0.00104	0.0140**	0.00892	0.00705	0.00473
	(0.00462)	(0.00430)	(0.00489)	(0.00485)	(0.00486)	(0.00596)
Used French as Child	1.078***	0.997***	1.002***	0.845***	0.822***	0.278*
	(0.153)	(0.124)	(0.163)	(0.129)	(0.133)	(0.115)
(French as Child)×Tenure	-0.0160**	-0.0157**	-0.0144*	-0.0129*	-0.0121*	-0.00523
	(0.00590)	(0.00490)	(0.00659)	(0.00542)	(0.00564)	(0.00471)
Variance						
Tenure	-0.0130***	-0.0152***	-0.0111***	-0.0119***	-0.0122***	-0.00699**
	(0.00326)	(0.00262)	(0.00323)	(0.00255)	(0.00254)	(0.00253)
Female	0.0632	-0.0125	0.114	0.00190	0.00405	0.0297
	(0.0682)	(0.0679)	(0.0699)	(0.0699)	(0.0702)	(0.0726)
Female×Tenure	-0.00173	-0.00111	-0.00299	-0.00265	-0.00216	-0.00211
	(0.00326)	(0.00303)	(0.00318)	(0.00300)	(0.00301)	(0.00293)
Refugee	-0.213	-0.127	-0.374**	-0.319**	-0.324**	-0.332*
	(0.122)	(0.119)	(0.130)	(0.116)	(0.117)	(0.129)
Refugee×Tenure	0.0122*	0.00830	0.0158**	0.0163**	0.0157**	0.0136*
	(0.00595)	(0.00598)	(0.00604)	(0.00555)	(0.00570)	(0.00586)
Used French as Child	0.155	0.0109	0.151	0.00294	-0.0111	0.0690
	(0.126)	(0.121)	(0.139)	(0.127)	(0.129)	(0.108)
(French as Child)×Tenure	0.00271	0.00754	0.00231	0.00719	0.00780	0.000251
	(0.00547)	(0.00510)	(0.00584)	(0.00529)	(0.00538)	(0.00458)
Education	No	Yes	No	Yes	Yes	Yes
Birthplace	No	No	Yes	Yes	Yes	Yes
Ling. Proxim.	No	No	No	No	Yes	Yes
Fluency at Migration	No	No	No	No	No	Yes
N	7601	7576	7601	7576	7547	7282

Note: Regressions are similar to Table 2.2, Panel (c).
* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$

2.3.4 Prediction 4: Un-Assimilation Should Occur, as Well

As a second implication of the addiction mechanism, the model predicts that immigrants should simultaneously un-assimilate to their mother tongue as they learn a new language. The Trajectories and Origins data is the only survey that asks about fluency in one's mother tongue, and Table 2.4 shows the results of a location-scale regression on mother tongue fluency. Since those who migrated very early in life may never have developed full proficiency in their native language, the regressions include only those immigrants who arrived after 15 years of age. Comparing Tables 2.2 Panel (c) and 2.4, the signs of every major variable are opposite for the coefficients on the mean, indicating that assimilation and unassimilation are mirror-image processes. The effect of age at migration is strong, but the effect of tenure is weak, implying that language attrition is not as evident as language learning. Tenure has a large effect on the variance, however, and variance tends to widen over time. This is consistent with mother tongue fluency forming a bimodal pattern for opposite reasons that we observe it for French.

Table 2.4 is important for the conceptual argument made by the model: language stocks L and M are treated as mutually exclusive, with investment in one only possible at the cost of depreciating the other. Since investments are made only through time spent using a language, disuse can be associated with "forgetting." Table 2.4 provides the only possible evidence from any of the four surveys that this conceptual framework is not terribly unrealistic: it appears that French and mother tongues are indeed substitutes.

Table 2.4: Location-Scale Probit: Mother Tongue Fluency (TeO)

	(1)	(2)	(3)	(4)	(5)
Mean					
Age at Migration (AaM)	0.131***	-0.0410	0.168***	0.0143	0.161***
	(0.0272)	(0.0272)	(0.0286)	(0.0284)	(0.0285)
$(\text{AaM})^2/100$	-0.205***	0.0634	-0.298***	-0.0523	-0.291***
	(0.0376)	(0.0391)	(0.0419)	(0.0419)	(0.0417)
Tenure	-0.0197	0.00533	-0.00941	0.0147	-0.0150
	(0.0291)	(0.0279)	(0.0260)	(0.0260)	(0.0256)
$\text{Tenure}^2/100$	0.0130	-0.0186	-0.0532	-0.0805*	-0.0560
	(0.0455)	(0.0455)	(0.0360)	(0.0390)	(0.0354)
(AaM)×Tenure	0.000472	0.000887	0.000430	0.000986	0.000679
	(0.000714)	(0.000646)	(0.000653)	(0.000641)	(0.000660)
Female	0.0335	-0.0720	-0.0851	0.167	-0.0919
	(0.254)	(0.224)	(0.206)	(0.223)	(0.204)
Female×Tenure	-0.0128	-0.00667	-0.0118	-0.0149	-0.0110
	(0.0110)	(0.00995)	(0.00849)	(0.00910)	(0.00834)
Variance					
Tenure	0.00820**	0.00807**	0.00781**	0.00912***	0.00797**
	(0.00283)	(0.00262)	(0.00253)	(0.00247)	(0.00258)
Female	0.0291	-0.152	0.0156	0.0250	-0.00783
	(0.120)	(0.109)	(0.121)	(0.114)	(0.122)
Female×Tenure	-0.00130	0.000963	-0.00358	-0.00465	-0.00309
	(0.00390)	(0.00356)	(0.00473)	(0.00435)	(0.00490)
Education	No	Yes	No	Yes	No
Birthplace	No	No	Yes	Yes	Yes
Ling. Proxim.	No	No	No	No	Yes
N	7574	7549	7574	7549	7567

Coefficients are from an ordered location-scale probit regression of a 4-category ordinal measure of fluency in one's mother tongue: . Standard errors are in parentheses.
SOURCE: Trajectories and Origins Survey (TeO). Regressions include all non-native French speakers, limited to those who arrived after the age of 15. Linguistic proximity to French was calculated according to method of Fearon (2003).

2.3.5 *Prediction 5: Assimilation Along Other Margins Should Be Similar*

The model is agnostic with respect to the type of cultural capital being used in social interactions. If this abstraction is warranted, then several other measures of assimilation should show the same patterns as linguistic fluency. A host of other behaviors, beliefs, and customs inform one's ability to interact with others; these aspects of daily life should manifest themselves as feeling that one "belongs to" or "identifies with" one's community.

I begin by examining assimilation secondary languages. One virtue of studying immigrants to Spain is that they can assimilate to Spanish as well as to any of Spain's three official minority languages (Catalan, Basque, and Galician). For those immigrants who moved to an applicable Spanish region (including those who are native Spanish speakers), we can estimate assimilation parameters for the minority languages. Table 2.5 shows the results, which follow the same pattern as for Spanish fluency, regardless of whether native Spanish-speakers are included or not. The table pools speakers of all three minority languages. The coefficients are not as large as in other regressions, but this is to be expected as the majority of individuals reported zero fluency in the secondary regional language. Hence, these regressions are somewhat confounded by selection problems, and future work could examine which immigrants have greater incentives to learn regional languages, and why. The table indicates, for a start, that women have lower levels of fluency, although the difference is small.

TeO is the only survey to ask about feelings of belonging, social identity, religious practice, and other aspects of social integration. Using the same regression specification as for language, I test whether addiction is a salient framework for describing changes in these behaviors over time, including assimilation towards "French-ness" and un-assimilation from one's birthplace.

Table 2.6 shows the results for identity assimilation and unassimilation; the survey questions ask whether one "Feels French" or "Feels like one's birthplace." The evidence is certainly not as strong as for linguistic assimilation, but "feeling French" increases over time, and weakly increases in variance. Age at migration is insignificant, possibly reflecting that one cannot be exposed to "Frenchness" prior to migration, or possibly reflecting a different underlying acculturation process for identity versus language use. Females in particular are less likely to feel French. The fifth column shows coefficients for un-assimilation from one's

Table 2.5: Location-Scale Probit: Regional Language Fluency (ENI)

	Non-Native Spanish Speakers	Full Sample	Non-Native Spanish Speakers	Full Sample	Non-Native Spanish Speakers	Ful Sample
Mean						
Age at Migration (AaM)	-0.00557	-0.0471***	-0.0172	-0.0594***	-0.0208	-0.0590***
	(0.0153)	(0.00966)	(0.0168)	(0.0105)	(0.0177)	(0.0108)
$(\text{AaM})^2/100$	-0.0468*	0.0204	-0.0364	0.0334**	-0.0329	0.0322*
	(0.0191)	(0.0112)	(0.0207)	(0.0122)	(0.0217)	(0.0127)
Tenure	0.0523***	0.0519***	0.0542**	0.0480***	0.0610**	0.0511***
	(0.0157)	(0.0119)	(0.0186)	(0.0127)	(0.0208)	(0.0142)
$\text{Tenure}^2/100$	-0.0931***	-0.0853***	-0.110**	-0.0860***	-0.111**	-0.0963***
	(0.0275)	(0.0177)	(0.0354)	(0.0200)	(0.0393)	(0.0231)
$(\text{AaM})\times\text{Tenure}$	0.0000560	-0.0000224	0.000158	0.000130	0.000234	0.000247
	(0.000321)	(0.000246)	(0.000363)	(0.000251)	(0.000373)	(0.000244)
Female	-0.000241	0.0278	-0.0370	-0.0305	-0.0612	-0.0409
	(0.145)	(0.0870)	(0.150)	(0.0907)	(0.148)	(0.0883)
$\text{Female}\times\text{Tenure}$	0.00311	-0.000912	0.00305	0.000298	0.00404	-0.000290
	(0.00759)	(0.00411)	(0.00836)	(0.00447)	(0.00850)	(0.00455)
Variance						
Tenure	0.00788	0.00856*	0.0126*	0.0116**	0.0137*	0.0123**
	(0.00586)	(0.00353)	(0.00612)	(0.00359)	(0.00618)	(0.00376)
Female	0.0200	0.0604	0.0251	0.108	-0.0000528	0.0467
	(0.126)	(0.0794)	(0.125)	(0.0792)	(0.125)	(0.0803)
$\text{Female}\times\text{Tenure}$	-0.00283	-0.00148	-0.00276	-0.00331	-0.00177	-0.00215
	(0.00662)	(0.00370)	(0.00645)	(0.00367)	(0.00624)	(0.00368)
Educ. on Arr.	No	No	Yes	Yes	Yes	Yes
Ling. Proxim.	No	No	No	No	Yes	Yes
N	3483	6532	3483	6532	3454	6503

Coefficients are from an ordered location-scale probit regression of a 5-category ordinal measure of Catalan/Galician/Basque fluency. Sample consists of individuals living in Spanish autonomous regions in which there is a second official language. Standard errors are in parentheses.

SOURCE: National Immigrant Survey (ENI). Linguistic proximity to the regional language was calculated according to method of Fearon (2003).

birthplace. Neither the signs nor the magnitudes are convincing.

A more precise story can be told by performing group-specific regressions. African and Muslim immigrants are plentiful in France but possibly face different structural barriers to assimilation compared to Europeans, due to skin color, religion, or other characteristics. The second, third, and fourth columns of Table 2.6 show the coefficients for each group. The coefficients for mean fluency generally have signs that are consistent with the linguistic assimilation patterns shown above, but only for Muslims are they particularly strong. The sixth, seventh, and eighth columns show that un-assimilation does not work the same way; Sub-Saharan Africans, if anyone, appear to give up their identity more strongly than other groups.

The different patterns of identity assimilation (or lack thereof) are not due to differing levels of French identity at arrival: Muslims with tenure less than 5 years actually report higher average levels of French-ness than Europeans or Africans, and the regression indicates those with longer tenure have even higher levels of identity. But given the subjective nature of the survey questions, the patterns could be interpreted in multiple ways. For example, the apparent lack of European immigrants' identity assimilation could be because they are already culturally similar enough to France that their is no need adapt identities. It could also be that Europeans interpret the survey question differently, having different standards for what qualifies as "French." As for Muslims, their apparent tendency to assimilate may reflect flies in the face of much media coverage of French immigrant communities. But it may also be the result of older immigrants revising their standard for what qualifies as "French." This is one reason a more comprehensive index of assimilation would be important for drawing conclusions.

Religious assimilation is another possible margin to consider. Assimilation here means Finally, Table 2.7 displays results for religious identity: do people deem it important to practice religion in general, maintain a religious diet, or dress according to religious guidelines? In France, a generally secular Catholic country, "assimilation" would mean decreasing im-

Table 2.6: Location-Scale Probit: Identity Assimilation (TeO)

	Full Sample	Muslim	Sub-Saharan	European	Full Sample	Muslim	Sub-Saharan	European
Mean								
Age at Migration (AaM)	-0.0240	-0.0348*	-0.0176	-0.0445	-0.00597	0.0314	-0.0685**	0.0177
	(0.0160)	(0.0168)	(0.0244)	(0.0254)	(0.0167)	(0.0185)	(0.0224)	(0.0290)
$\text{AaM}^2/100$	0.0379	0.0881**	0.0219	0.0646	-0.0174	-0.0636*	0.0437	-0.0332
	(0.0255)	(0.0338)	(0.0413)	(0.0333)	(0.0234)	(0.0286)	(0.0335)	(0.0375)
Tenure	0.0401*	0.0512**	-0.0292	0.0576*	-0.0228	0.0217	-0.107**	-0.0347
	(0.0177)	(0.0197)	(0.0309)	(0.0260)	(0.0211)	(0.0253)	(0.0376)	(0.0249)
$\text{Tenure}^2/100$	-0.0203	-0.0576	0.0836**	-0.0262	-0.0419	-0.0613	0.0452	0.0385
	(0.0269)	(0.0320)	(0.0298)	(0.0340)	(0.0320)	(0.0420)	(0.0710)	(0.0284)
(AaM)×Tenure	-0.000295	-0.000368	0.000348	-0.000526	0.00139**	-0.0000727	0.00348***	0.000513
	(0.000400)	(0.000422)	(0.000617)	(0.000624)	(0.000471)	(0.000445)	(0.000787)	(0.000619)
Female	-0.245**	-0.282**	-0.422**	-0.206	0.123	0.0917	-0.0917	-0.251
	(0.0935)	(0.103)	(0.159)	(0.128)	(0.116)	(0.165)	(0.258)	(0.161)
Female×Tenure	0.00928	0.00532	0.0244*	0.0112	-0.0132*	-0.00629	0.0123	0.000571
	(0.00602)	(0.00523)	(0.0122)	(0.00658)	(0.00590)	(0.00678)	(0.0175)	(0.00573)
Variance								
Tenure	0.000317	0.00675	-0.0110	0.00530	0.00428	-0.00292	-0.0136	-0.00927
	(0.00297)	(0.00363)	(0.0231)	(0.00303)	(0.00347)	(0.00353)	(0.00885)	(0.00552)
Female	-0.0281	-0.0537	-0.114	-0.0241	-0.201	-0.269	-0.404	-0.328
	(0.0713)	(0.105)	(0.342)	(0.117)	(0.107)	(0.160)	(0.316)	(0.206)
Female×Tenure	0.00233	-0.00148	0.00724	0.00327	0.00883	0.00830	0.0348	0.00823
	(0.00340)	(0.00458)	(0.0251)	(0.00427)	(0.00504)	(0.00611)	(0.0210)	(0.00633)
N	8456	5555	1283	2345	8573	3271	1320	2346

Note: Coefficients are from an ordered location-scale probit regression, where answers are on a 4-category ordinal scale.
SOURCE: Trajectories and Origins Survey (TeO). Regressions include all non-native French speakers.
* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$

portance of religion, particularly for Muslim immigrants. The results here are more mixed, but in general people do place less importance on religion the longer they stay in France. The variance parameters show an especially mixed response between diet and clothing, although the decision to wear religious clothing seems to imply greater heterogeneity over time. Interestingly, age at migration has negative coefficients, possibly showing that older immigrants tend to be less religious in the first place.

2.4 Conclusion

This chapter offered a variety of evidence for the increasing-heterogeneity state of the world predicted in Chapter 1. Using the capital-accumulation-and-depreciation framework of the model, many more predictions can be made. Economic outcomes of assimilation could be examined, and many more "price experiments" like those in this paper could be conducted. For instance, different sex ratios between immigrant groups imply different incentives to assimilate, in order to marry – but assimilating for the marriage market can result in different economic or occupational outcomes as well. Capital assets could also be studied: immigrant groups with higher incentives to remit money (such as those from disaster-struck countries) also have incentives to maintain a connection to their birthplace, meaning they are both more capital-constrained and less socially-assimilated. Structural models could help distinguish the different effects on both of these margins.

Due to their small size but rich detail, the surveys studied in this paper would lend themselves well to further structural modeling, using the blueprint presented here as a way to think about trajectories and complementarity/substitutability in time investments.

Table 2.7: Location-Scale Probit: Religious Identity (TeO)

	Overall Religiosity		Religious Diet		Religious Clothing	
	Full Sample	Muslims	Full Sample	Muslims	Full Sample	Muslims
Mean						
Age at Migration (AaM)	-0.0105	-0.0180	-0.0523**	-0.105***	0.00561	0.0570*
	(0.0226)	(0.0193)	(0.0182)	(0.0278)	(0.0240)	(0.0262)
$AaM^2/100$	-0.00161	0.00107	0.0605*	0.151***	0.00107	-0.0911
	(0.0307)	(0.0286)	(0.0257)	(0.0442)	(0.0309)	(0.0481)
Tenure	-0.0243	-0.0378	-0.0467***	-0.0416	-0.0533***	-0.0184
	(0.0256)	(0.0210)	(0.0113)	(0.0291)	(0.0108)	(0.0258)
$Tenure^2/100$	-0.00281	0.0440	0.0254	0.0567	0.0579***	0.0155
	(0.0390)	(0.0322)	(0.0169)	(0.0407)	(0.0157)	(0.0340)
(AaM)×(Tenure)	0.00109	0.00110*	0.00130**	0.00214***	-0.000259	-0.000920*
	(0.000560)	(0.000456)	(0.000412)	(0.000618)	(0.000531)	(0.000448)
Female	0.238	0.310*	-0.116	0.676	-0.611	4692.4***
	(0.134)	(0.127)	(0.109)	(0.480)	(1.039)	(1227.0)
Female×Tenure	-0.00312	-0.00504	0.00325	-0.0301	0.00530	122.1*
	(0.00730)	(0.00521)	(0.00368)	(0.0229)	(0.00823)	(59.66)
Variance						
Tenure	0.00507	-0.000609	-0.0250**	0.00807	-0.0337	-0.0305
	(0.00420)	(0.00325)	(0.00807)	(0.00964)	(0.0272)	(0.0225)
Female	0.0225	-0.0784	-0.248	0.0326	-0.312	7.969***
	(0.105)	(0.122)	(0.256)	(0.323)	(0.587)	(0.208)
Female×Tenure	-0.00566	-0.00174	0.00120	-0.0105	-0.0248	0.0417
	(0.00573)	(0.00480)	(0.0106)	(0.0147)	(0.0435)	(0.0243)
N	7006	3282	5617	3282	8728	3304

Note: Coefficients are from an ordered location-scale probit regression. The outcome variable for general religiosity is on a 4-category scale; diet has 3 categories; and clothing is binary.
SOURCE: Trajectories and Origins Survey (TeO). Regressions include all non-native French speakers.
* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$

CHAPTER 3

RE-EVALUATING THE RETURNS TO LANGUAGE SKILLS USING LATENT FACTOR ESTIMATES

3.1 Introduction

Language skills are an important component of immigrants' human capital, correlating with immigrants' wages (McManus *et al.* , 1983; Chiswick, 1991; Borjas, 2014), propensity for return migration (Dustmann & Görlach, 2016), neighborhood sorting (Lazear, 1999), employment status (Dustmann & Fabbri, 2003), occupational choice (Schoellman, 2010), and skill complementarity (Berman *et al.* , 2003). Measuring all of these outcomes is, in turn, vital for understanding the full economic impact of immigration on receiving countries. Nevertheless, it is well-established that the raw data used to measure language skills – typical self-reported fluency on an ordinal Likert scale – can be unreliable for several reasons, including misclassification bias, recall bias in retrospective questions, or systematic differences.

This paper uses standard statistical methods to re-evaluate the results of some seminal papers, using latent indices of language skill. These indices account for multidimensionality in language skills and systematic differences in response patterns across groups and across survey questions. Past research focused on correcting for misclassification bias in fluency data, for instance when respondents over- or under-estimate their fluency. This is especially problematic with self-reported survey questions (Hausman *et al.* , 1998), and Dustmann & van Soest (2001, 2004) use latent factor methods to correct the bias in both panel and cross-sectional data. But correcting for misclassification error may be insufficient for two reasons, both illustrated by Table 3.1 using data from the New Immigrant Survey (one of the two surveys used in this paper). First, Panel 3.1a shows that using different measures of fluency may lead to different estimates of underlying latent skill, since respondents' own evaluation of their English skills may contradict the interviewers' evaluations. Most studies use the former measure, while Dustmann & van Soest (2004) use the latter. Which is more appropriate? Ideally, a latent factor should account for misclassification bias both within and across the two questions. Second, Panel 3.1b shows that even without misclassification bias, a single measure of fluency can be mis-informative. The panel shows that respondents' own evaluation of their skills in speaking and understanding do not perfectly align. This does not imply misclassification: people can indeed have different skills in one versus the other. But how should researchers incorporate the information from both variables?

Table 3.1: Inconsistent Answers in the TeO Survey

(a) Two Assessments of Respondent's French Comprehension

Interviewer	Self: 1 (Not Well)	2	3	4 (Very Well)
1 (Not Well)	32	299	67	17
2	1	302	647	161
3 (Well)	3	76	776	2996

(b) Self-Assessment: Speaking vs. Understanding

Understand	Speak: 1 (Not Well)	2	3	4 (Very Well)
1 (Not Well)	29	5	2	0
2	17	578	83	3
3	1	186	1231	95
4 (Very Well)	0	17	281	2934

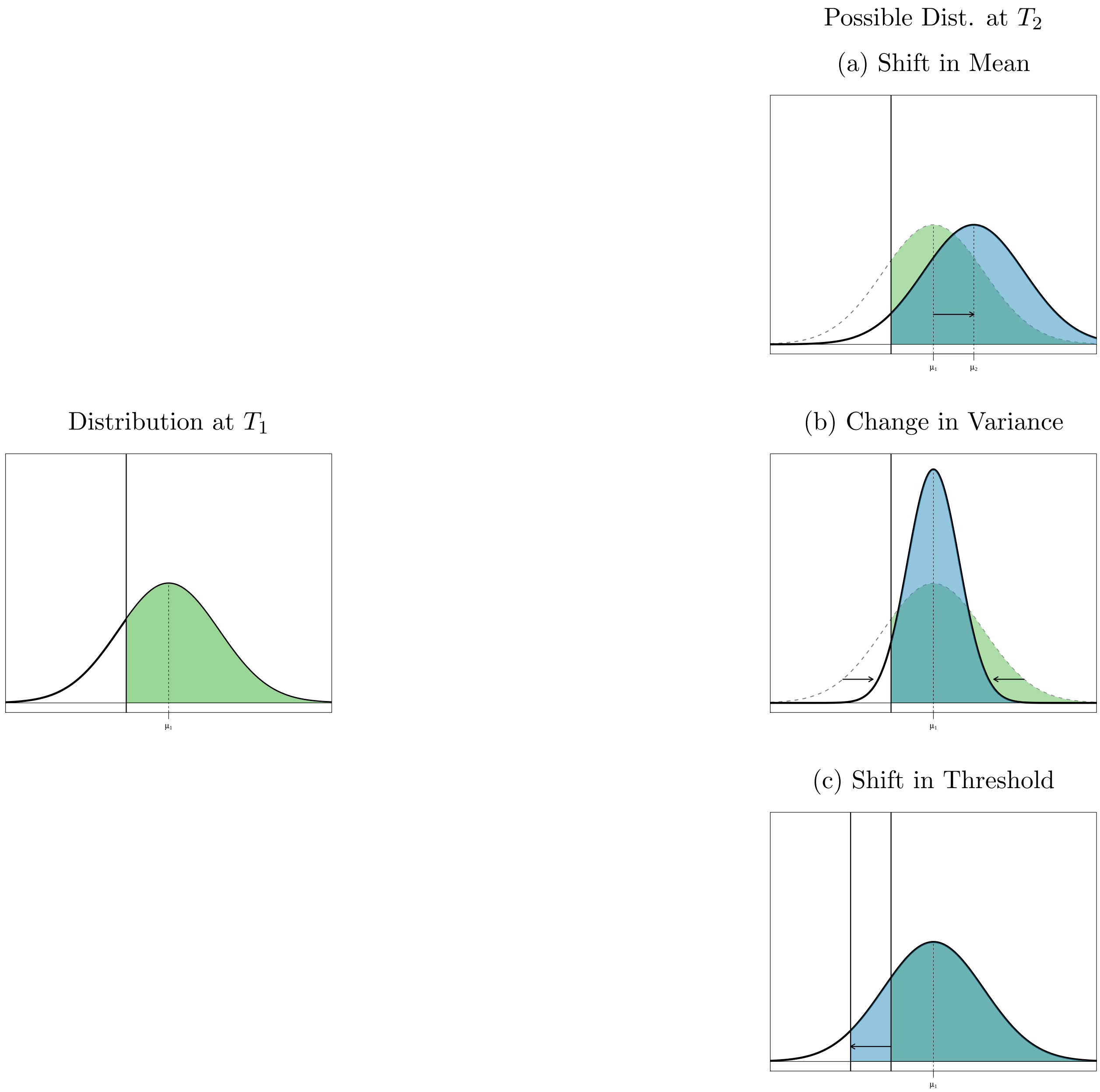
Note: Each panel shows frequency counts for cross-tabulations of different measures of French language skill.

The current approaches to multidimensionality are relatively ad-hoc. Some studies take means over different measures of fluency, some use multiple separate measures (Dustmann, 1994), and some sort respondents into discrete groups based on combinations of survey answers (McManus, 1985). It is not obvious that any given approach is superior to any other. At any rate, the resulting discrete variables impede inference anyway. When the distribution of responses is of interest (as opposed to the mean or a binary summary of the data), changes in the distribution can be mis-identified or misinterpreted, as illustrated by Figure 3.1. Suppose a survey asks "How well do you speak English?" with two possible answers: "Well" or "Not well." The left panel shows the initial continuous distribution of fluency, with the threshold above which people answer "Well." The conclusion from the data would be that 80% of people speak well. The right panels show three possible changes at the time of a second survey, including a change in the perceived threshold as to what constitutes "speaking well." Each of these new distributions leads to the conclusion that 95% of people speak well. However, the changes in fluency for inframarginal respondents, as well as the reasons behind the changing mean, are vastly different and cannot be distinguished, even though such inframarginal responses may be of interest to researchers.

The latent measures of language skill developed in this paper are calculated using the graded response model (GRM), a standard method from the psychometric literature on extracting information from responses to test questions with ordinal answer scales (Samejima, 1969; Muraki & Carlson, 1995). The latent measures are also robust to misclassification caused by systematically different response patterns between groups. This form of misclassification, called differential item functioning (DIF), occurs when two respondents in different demographic groups but with the same latent language skill give different answers to the same question (see Zumbo, 2007, for a historical overview of DIF detection methods). DIF may be caused, for instance, because respondents with the same grammatical skill and vocabulary have different reference points for what it means to speak English "well." Someone who uses English at work may believe their English to be "worse" than it "really" is, while someone with the same vocabulary and grammar skill who uses their native language at work may believe their English to be "better." This will result in systematic differences in responses to the question "How well do you speak English?"

The latent factors are calculated using detailed sociological survey data from the Trajectories and Origins survey (TeO), a nationally representative survey of immigrants in France.

Figure 3.1: Lack of Identification for Social Variables.



Note: Left panel shows hypothetical distribution of language fluency at initial survey date T_1. 80% of the sample is above the threshold. Right panels show alternative ways in which the distribution or threshold could shift at second survey date T_2, all of which yield 95% of the distribution above the threshold.

The virtue of using such a detailed dataset is that it provides multiple measures of language fluency, and latent factors can be calculated in multiple ways to determine the robustness of the measures. Using the resulting measures of latent language skills, I replicate the results from seminal papers measuring the effect of language skills on economic outcomes, including wage premia, labor force participation, and marriage. Section 3.2 describes the statistical methods used for calculating latent factors via GRM and adjusting for differential item functioning, Section 3.3 describes the data, and Section 3.4 outlines the results of the latent variable estimation. Section 3.5 describes the results of the replication exercises, and Section 3.6 concludes.

3.2 Method

3.2.1 Constructing Latent Factors: Graded Response Model

As with any latent factor analysis of language skills, the basic assumption for this exercise is that a person's linguistic human capital stock L is unobserved, but manifests certain observable skills and behaviors. The measures of those skills and behaviors must be used to extrapolate the factor.

Survey data on language skills and language use are ideal inputs into the graded response model (GRM), developed originally for binary data by Samejima (1969) and expanded to polytomous ordinal data by Muraki & Carlson (1995). Because GRM comes from the psychometrics literature and may be unfamiliar, this section briefly reviews it. Further details can be found in Reckase (2009, §2.1.3.3 and §4.1.2.3).

The components of the graded response model are survey questions $q \in \{1, \ldots, Q\}$, each of which has possible responses $r \in \left\{1, \ldots, R_q\right\}$. Responses are hierarchical. In other words, response r also implies or subsumes response $r-1$. Likert scales fit this framework perfectly, since a choice of one rating category necessarily subsumes or outranks all lower categories.

The data for each individual n consists of responses $\left\{x_{qn}\right\}_{q=1}^{Q}$. The probability of rating at least r on question q is $\mathrm{P}_q\left[r\middle|L\right]$. The probability of scoring exactly r is $\mathrm{p}_q\left[r\middle|L\right] = \mathrm{P}_q\left[r\middle|L\right] - \mathrm{P}_q\left[r-1\middle|L\right]$, which can be plotted as a function of L to give category response curves. I adopt the standard framework and use the logistic form of the GRM, so that the probabilities are given by equation 3.1. The question-specific discrimination parameter γ_q determines how well question q distinguishes different latent skills L, and alters the width of the category response curves. The response-specific difficulty parameter β_{qr} determines the latent skill L that is most likely to answer the question correctly, thereby shifting the peak of the category response curve right or left.

$$\mathrm{P}_q\left[x_{qn} \geq r\middle|L_n\right] = \begin{cases} 1 & r = 0 \\ \frac{1}{1+\exp\left(-\gamma_q\left(L_n - \beta_{qr}\right)\right)} & 0 < r < R_q \\ 0 & r = R_q \end{cases} \tag{3.1}$$

Assuming the distribution of latent traits is $f(\cdot)$, the maximum likelihood contribution of respondent n is equal to the the probability of n's vector of responses (equation 3.2).

$$\mathcal{L}_n\left(\boldsymbol{\gamma}, \boldsymbol{\beta}\right) = \int \mathrm{p}\left[\boldsymbol{x}_n\middle|\boldsymbol{\gamma}, \boldsymbol{\beta}, L_n\right] f\left(L_n\right) dL_n \tag{3.2}$$

3.2.2 Accounting for Group Differences: Differential Item Functioning

Differential item functioning (DIF) refers to a difference between groups in their response pattern on a question or set of questions, controlling for latent skill. It is important to match respondents' latent skills correctly, so that real differences between groups are not misidentified as DIF. However, matching is not a trivial exercise. When DIF is present then any calculation of the latent trait will be "impure" and cannot necessarily provide a valid match. Yet some sort of matching criteria must be used in order to identify DIF in the first place. To solve this conundrum, several "purification" options have been proposed, each with their own limitations. I use the iterative method of Crane *et al.* (2006), which estimates latent traits on the full sample, then identifies the questions with DIF. It then estimates GRM parameters separately for questions found to have DIF (group-specific estimates) versus those that do not (full sample estimates); using these newly-estimated parameters the DIF questions are re-flagged. The procedure is repeated until the same set of questions is flagged. This approach preserves the information from all questions, rather than discarding those questions with DIF, and also avoids misidentification of DIF in early stages of the loop.

Given an estimate of respondents' latent traits, the standard method for DIF detection is a χ^2 likelihood ratio test (Swaminathan & Rogers, 1990). Cumulative logit models determine whether the probability of different responses depends on the latent trait only (model 1), the trait and the group membership (model 2), or an interaction between the two (model 3). A likelihood ratio test between models 1 and 2 identifies uniform DIF (where the differences between groups are constant at all levels of the trait) and a test between models 2 and 3 identifies non-uniform diff (where the difference depends trait level). The method of Choi *et al.* (2011), implemented in R package `lordif`, performs the method just described and yields DIF-free estimates of the latent trait L_n.

3.3 Data

Robust measures of language fluency require data with more variables than are found in many large surveys. The patterns documented by Table 3.1 can only be characterized when data is sufficiently rich as to allow for multiple measures of language skill and use. Ideally, survey data will allow for corroboration of responses using multiple forms of evidence (to account for misclassification of the sort suggested by Panel 3.1a) and for measurement of skill along multiple margins (to account for multidimensionality of the sort suggested by Panel 3.1b). Unfortunately, many commonly-used surveys, such as the U.S. Census and the American Community Survey, fail to satisfy both conditions.

Data must be also be rich in variation in order to run the GRM and DIF algorithms. Data must be split into DIF-relevant cells, and questions' response vectors must not be too correlated within any given cell. In addition, in order to be informative, each question must have sufficient variation across responses within each cell. The need for variation puts limitations on the questions used to identify the latent factor and the construction of cells for DIF adjustment.

To counter the data problems of many surveys, I use data from the Trajectories and Origins survey (TeO), a French sociological survey that offers detailed information on immigrants' entire migration history, employment history, education, marriage, family structure, assets, loans to and from other family members, refugee status, remittances, and neighborhood demographics. TeO is nationally-representative of all immigrants in France (see Beauchemin & Simon, n.d., for the project description). Complete details for constructing variables may be found in the online data dictionary.

The Trajectories and Origins Survey (*Trajectoires et Origins*, or TeO) encompasses 21,761 respondents. The survey took place in 2008 and 2009 with a stratified sampling strategy to capture a representative sample from five different segments of the French population: immigrants (N=8734), children of immigrants (8241), migrants from French overseas departments (*départements d'outre-mer* or DOMs, N=725), children of DOM-ers (653), and children of natives (3408). I use non-native French speakers from the immigrant sample; in accordance with previous studies, I consider only those respondents who emigrated at age 18 or later. This leaves 5031 respondents.

Table 3.2 provides summary statistics on this sample, and provides a comparison between respondents who reported speaking well (4 on the 1-4 scale of language skill) versus those who did not. Unsurprisingly, speaking well correlates with speaking some French as a child; having better education; marrying someone from outside one's birthplace; being male, non-refugee, and employed; arriving earlier in life; and having a longer tenure in France. Sub-Saharan immigrants are more likely to speak well, reflecting the linguistic heritage of certain Africa countries as well as likely selection in immigration flows. However, there not substantial differences in the distribution of age at migration and tenure, between those who speak well and those who do not.

Table 3.2: Summary Statistics for TeO Survey

	Full Sample	Speak Well	Do Not Speak Well
N	5031	2587	2444
French Skills			
Ever Spoke as Child (%)	16.2	24.8	7.1
Spoke Well on Arrival (%)	24.8	49.2	0
Age at Migration			
Mean	27	26	28
Median	25	24	26
Max	59	55	59
Tenure in Country			
Mean	16.6	17.3	15.8
Median	15	16	13
Max	42	42	42
Demographics			
European (%)	23.8	23.5	24.1
MENA Region (%)	36.6	35.6	37.6
Sub-Saharan (%)	21.7	28.6	14.4
Female (%)	54.2	52.0	56.5
Refugee (%)	13.4	9.6	17.4
Married (%)	75.3	71.1	79.8
Endogamous (%)	52.7	41.8	64.3
Has Children (%)	83.4	81.3	85.7
High Local Imm. Density (%)	74.2	71.7	76.7
Education & Employment			
H.S./2-yr Coll. (%)	18.2	22.3	14.0
4-yr College or More	22.0	32.6	10.8
Employed (%)	63.7	71.4	55.6
In School (%)	3.5	2.7	4.3

Note: Sample consists of immigrants who arrived at age 18 or later and do not speak French as a native language. Variables are described in Section 3.3. See online data dictionary for complete details on constructing variables from raw survey data.

To measure neighborhood effects on language skills, I use the recorded density of immigrants in one's neighborhood. The TeO survey reports density in terms of ventiles, based on neighborhoods across the whole of France. Most neighborhoods have zero or very few immigrants; thus, among neighborhoods where immigrants actually live, the variation in ventiles is quite low. I therefore collapse the reported ventiles into a binary variable indicating whether the respondent's neighborhood has more than a median concentration of immigrants, compared to the country as a whole. Over 70% of respondents live in one of these neighborhoods.

TeO offers a wide variety of questions measuring language skills. The possible questions are as follows:

1. "How well do you speak French?" (1-4 scale)
2. "How well do you understand French?" (1-4 scale)
3. "How well do you read French?" (1-4 scale)
4. "How well do you write French?" (1-4 scale)
5. Interviewer's assessment of respondent's French (1-3 scale)
6. To what extent was the interview translated? (1-3 scale)
7. "Do you have difficulty giving your name and phone number in French?" (binary)
8. "Do you have difficulty answering simple questions in French?" (binary)
9. "Do you have difficulty asking for information or services in French?" (binary)
10. "Do you have difficulty taking part in a conversation in French?" (binary)
11. "Do you have difficulty describing something in detail in French?" (binary)

The paths through the survey are somewhat complicated, and result in many questions being asked only of certain sub-populations. For example, those who report speaking "very well" were not subsequently asked if they had difficulty asking for information or services. Answers are filled in accordingly when the questions were not asked (details available in online appendix).

As further evidence for the insufficiency of unidimensional measures of language fluency, Figure 3.2 plots the implied density of skills based on responses to questions 1-4. The distribution varies for different skills, and while most appear to be skewed bell-shaped curves, the reported skill in writing is bimodal. Collapsing these measures into binary variables may imperfectly characterize the heavy left tail as being more homogeneous than it really is; and using any one of the four measures omits information about complementarities and substitutabilities between different skills. The latent factor estimates can help determine the degree to which different skills matter.

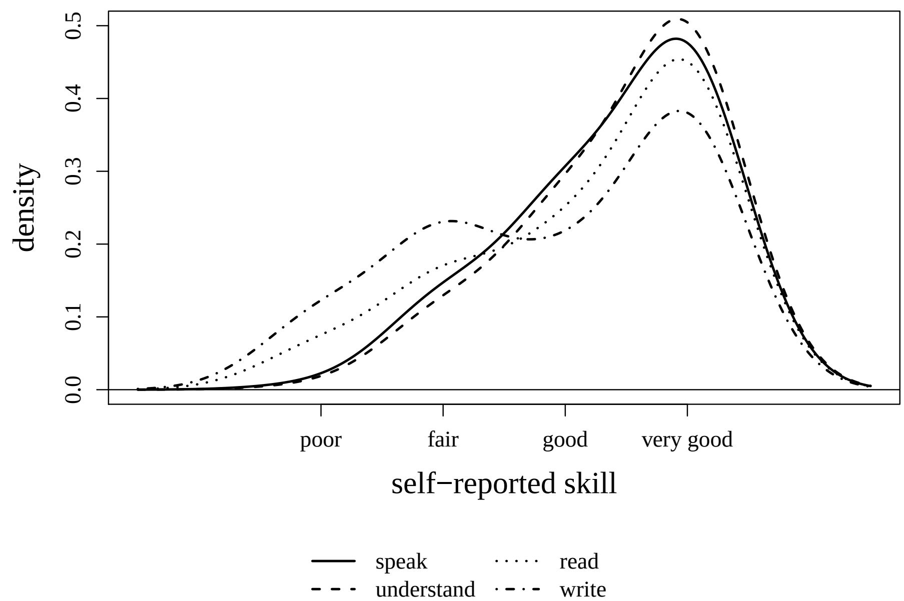


Figure 3.2: Implied Distributions of French Skills

Note: Densities are calculated from four survey questions, each with four possible response categories.

3.4 Latent Factor Estimation Results

The baseline latent factor estimate is calculated from equations 3.1 and 3.2 using all 11 questions listed above. It is useful to compare the baseline results to the commonly-used binary measure of "speaking well." Figure 3.3 plots the density of the estimated latent factor for individuals who report speaking well versus those who do not. It is immediately apparent that the binary measure of fluency implies a very different distribution from that given by the latent trait estimates. Among those who report speaking well, the average is $L = 0.67$, but the distribution is actually bimodal, a pattern strongly predicted by the theoretical model of Marrone (2017). This is further evidence of what was already shown in Tables 3.1a and 3.1b, in terms of speaking ability capturing only a partial picture of complete language skills.

Based on the distribution of latent traits, it is convenient to break up the population into three groups, labeled accordingly on Figure 3.3: those who speak poorly (Region A); those who speak well but in the left tail of the relevant distribution; and those who speak well in the right tail of the distribution. The means and modes are listed below:

1. Region A: mean -0.78, mode -0.29

2. Region B: mean 0.09, mode 0.40

3. Region C: mean 0.85, mode 0.85

Region B consists of respondents who report speaking well but who may have difficulty - to a greater or lesser degree - in speaking, reading, writing, etc. Region C consists of respondents who claim to speak well, but also report high ability to read, write, and comprehend, with little difficulty on any of the tasks surveyed, and with an interviewer assessment that corroborates these abilities.

Figure 3.3: Latent factor densities, by speaking ability

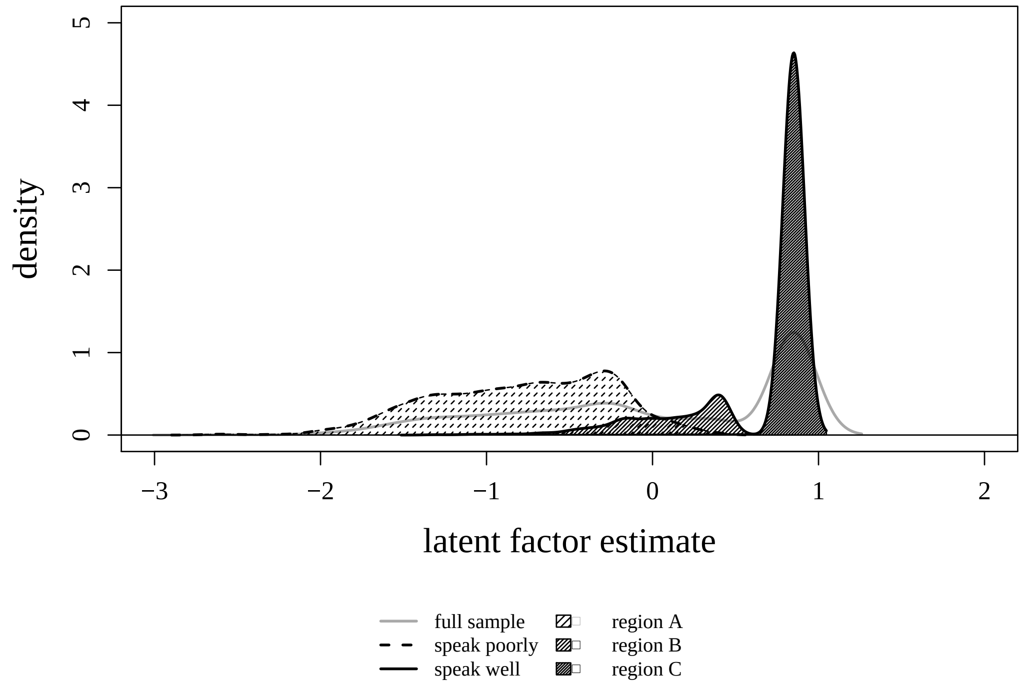


Figure shows probability density functions for baseline estimate of the latent language skill L_n, based on values of the binary variable "speak well."

Based on this observation, discrete measures of language ability will lead to misleading characterizations of the population's skills. The regression coefficient β on a binary variable such as "speak well" will characterize the difference in means between the two groups: $\hat{\beta} = \left(\bar{y}\middle|\text{speak well}\right) - \left(\bar{y}\middle|\text{speak poorly}\right)$. But average outcomes may not be a useful summary statistic when the dummy variable is summarizing a bimodal distribution. The most relevant comparisons would be between Regions A and B (the impact of speaking well, but not reading, etc.), or between B and C (the impact in reading, writing, and comprehending in addition to speaking). Therefore, even though a binary regressor has any easy interpretation and may accurately captures the difference in average respondents, it may obscure more useful patterns. The latent factor, in contrast, is difficult to interpret directly but allows for a more nuanced comparison of respondents with different combinations of skills. In the regressions of Section 3.5, I analyze when and how the binary measures of fluency fail to capture the same patterns as the continuous latent measures.

The baseline estimates shown in Figure 3.3 do not account for DIF. To do so, I first estimate latent traits L_n on the whole sample, then separately after detecting for DIF relative to various demographic characteristics. The possible DIF groupings include age at immigration, tenure in France, birthplace, education, gender, employment status, marital status, propensity to remit money, refugee status, and the fraction of immigrants in the respondent's neighborhood. Any of these variables may feasibly affect a person's view of what it means to "speak well" because they are evidence of different references or frames for the question. These frames result from different amounts of contact with native French speakers, as well as (possibly) cultural institutions and norms. The goal is to determine which DIF-detection groupings matter the most, in terms of their effect the latent factor estimates.

Variables are flagged for DIF at the 1% significance level, a number small enough to avoid flagging all variables, all the time. For each grouping variable, Table 3.3 marks the questions flagged for DIF (according to List X), as well as the correlation between the raw and purified estimates. For most variables, DIF was detected for some but not all of the 11 questions. In these cases, DIF purification does not necessarily result in economically relevant changes in the underlying latent factor estimates. The correlation between the raw and purified estimates is very high, and density plots show only small differences in the distribution of estimates.[1] In most cases, the individual DIF effects were small in magnitude but statistically significant, while in some cases the DIF on different questions offset each other, so that the overall effect was small.

1. Density plots for all variables are on file with the author.

Table 3.3: DIF Flagging, by Grouping Variable

	Question Number											
	1	2	3	4	5	6	7	8	9	10	11	Corr. (%)
Age at Migration	X		X	X	X	X	X			X		99.98
Birthplace	X	X	X	X	X	X	X	X	X	X	X	N/A
MENA vs. All Others	X	X	X	X	X	X						99.91
Sub-Sahara vs. All Others			X	X	X	X						99.89
Europe vs. All Others	X	X	X	X	X	X				X		99.81
"Other" vs. All Others	X	X	X	X	X	X	X	X	X	X	X	N/A
Education	X	X	X	X	X	X	X	X	X	X	X	N/A
<H.S. vs. All Others	X	X	X	X	X	X	X	X	X	X	X	N/A
H.S./2-yr Coll. vs. All Others	X	X	X	X	X							99.77
4-yr. College+ vs. All Others	X	X	X	X	X	X	X	X	X	X	X	N/A
Employed				X	X	X	X			X		99.98
Family Visa					X	X	X		X	X		> 99.99
Gender	X				X	X	X			X		> 99.99
Local Immig. Density												
All Immigrants			X	X	X							> 99.99
MENA		X	X	X	X	X	X			X		> 99.99
Sub-Saharan				X	X		X	X	X	X	X	99.97
Married					X	X						> 99.99
Refugee					X							> 99.99
Religion	X	X	X	X	X	X						99.74
Remittance in Last Year							X			X		> 99.99
Tenure	X	X	X	X	X	X	X	X	X	X	X	N/A
0-5 yrs vs. All Others	X	X	X	X	X	X	X	X	X	X	X	N/A
6-10 yrs vs. All Others				X	X	X						> 99.99
11-20 yrs vs. All Others												100
> 20 yrs. vs. All Others	X	X	X	X	X	X						99.96

Note: Table shows which variables were flagged for differential item functioning, based on various demographic grouping variables, and the correlation between raw and purified latent factor estimates.

But for birthplace, education, and tenure, DIF was detected in every question, meaning that a purified estimate could not be calculated because no questions could serve as anchors. This is one indication that these variables are possibly the most important to account for when estimating the latent factors. A second indication is that other groupings with a large number of flags, such as age at immigration and religion, are strongly correlated with tenure and birthplace. Since it is not possible to purify directly using the questions available in the survey, alternative methods must be used. One option is to use binary rather than multinomial groupings, if a salient dichotomy can be established. The subcategories of Table 3.3 show that when compared to all other respondents, those from the Middle East/North

Africa, Sub-Saharan Africa, and Europe show DIF on some but not all questions. The same is true of those who have been in France more than 5 years, and those who have a high school or associates degree. However, some birthplace, tenure, and education categories still show DIF on every question. Thus, there is no obvious regrouping that will preserve important distinctions between groups while making purification possible.

Instead, I opt for a hierarchical purification strategy, by detecting DIF on one variable separately for each sub-group determined by another variable. For example, one could perform separate DIF purification relative to tenure for each separate birthplace bucket. I choose tenure and birthplace because these variables have been found to have the strongest correlation with economic and cultural assimilation indicators (for evidence see Borjas (2014) and Chapters 1 and 2 of this dissertation). Purifying on, e.g., tenure within birthplace groups is an alternative to purifying directly across (tenure)×(birthplace) cells, which in this sample would yield cells that are too small to make the algorithm feasible.

Table 3.4: DIF Flagging, by Grouping Hierarchies

	Question Number					
	1	2	3	4	5	6
By Tenure, within Birthplace Group						
MENA	X	X	X	X	X	
Sub-Saharan Africa					X	
Europe	X		X	X		
Other	X		X	X	X	X
By Birthplace, within Tenure Group						
0-5 yrs		X	X	X	X	X
6-10 yrs			X	X		
11-20 yrs				X	X	
21+ yrs		X		X	X	X

Note: Table shows which variables were flagged for differential item functioning, based on two different demographic grouping hierarchies.

An obvious question is whether such a procedure is valid. Under group-specific purification, individuals are matched against respondents from the same group, not from the entire sample. One may concoct examples wherein within-group differences may be larger or smaller than between-group differences, so that DIF will be flagged for each group separately but not for the sample as a whole, or for the whole but not for separate groups. Table 3.4 shows the DIF flags when individual tenure groups are purified on birthplace, and vice-versa. Because of the resulting small cell sizes, only questions 1-6 were used (dropping the binary "difficulty" questions) to ensure convergence. Nevertheless, the table shows that within any group, DIF was only flagged on a strict subset of the questions. This is reassuring evidence that within-group differences are smaller than across-group differences, and that estimating and purifying within a group controls for important between-group heterogeneity.

To further assess the impact of this purification strategy, Figure 3.4 compares the latent factor estimation five ways. All methods yield qualitatively similar distributions: bimodal, with low-trait types covering a wide spectrum of latent trait values, and high-trait types being more similar. The raw estimate from the GRM algorithm using all 11 questions (dark gray solid line) and only the first 6 (black solid line) are extremely similar. Thus, the last five questions are sufficiently correlated with the first six that their exclusion is not problematic. Estimating separately for each tenure/birthplace cell narrows the implied gap between low and high types, while maintaining a heavier left tail. Purifying on birthplace for each tenure group preserves the original distribution most closely, although slightly exacerbating the distance between the two modes. Purifying on tenure for each birthplace group yields two sub-groups within the high types, one just to the right of the other. In the next section, results using both the tenure-by-birthplace and birthplace-by-tenure purified estimates are compared to results using the original raw latent factor.

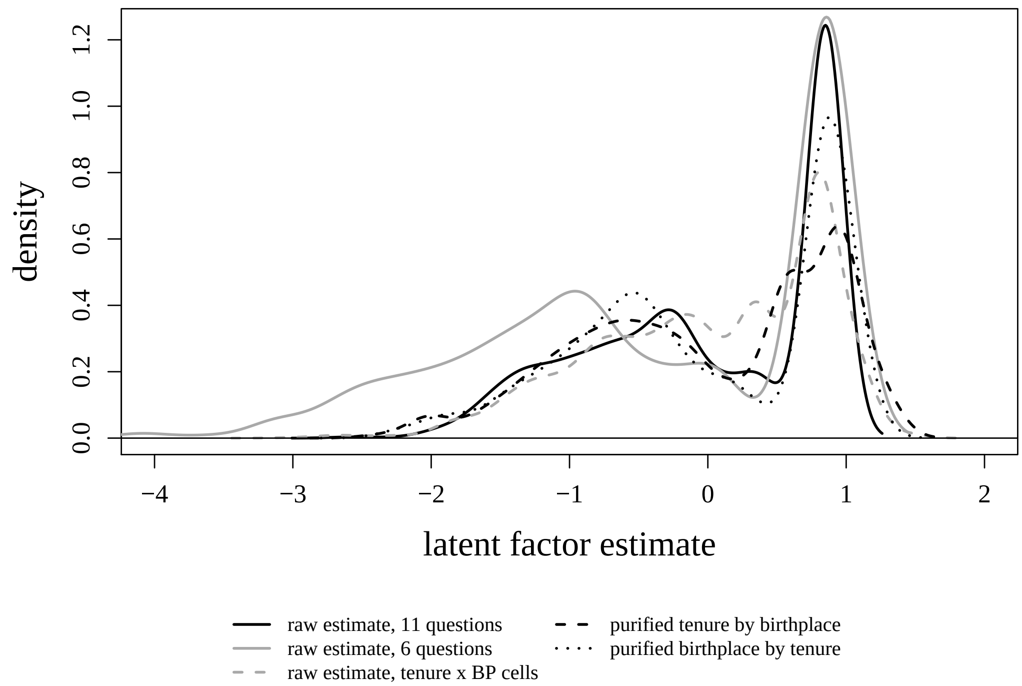


Figure 3.4: Distributions of Latent Factor Estimates

Note: The figure shows probability density functions for estimates of the latent language skill L_n, based on different combinations of questions and different hierarchical groupings for DIF purification.

3.5 Determinants and Effects of Language Skills

Using the purified latent trait estimates, I re-evaluate evidence from past studies on the economic effects of language skills. Previous studies have focused on two outcomes: the demographic determinants of language skills and the correlation between language skills and wages. I examine both in turn.

To assess the determinants of language skills, studies generally regress skills on demographic characteristics. Comparison across studies is difficult, because language skills are measured in slightly different ways and different regressors may be included, not to mention probit coefficients are difficult to interpret if not reported as marginal effects. Nevertheless, for results from other studies, see Chiswick (1991, Tables 3 and 5), Dustmann (1994, Table 3), Chiswick & Miller (1995, Table 1), Lazear (1999, Table 3), Dustmann & van Soest (2004, Tables 2 and 3). All studies agree that language skills increase with time since migration and decrease with age at migration. Married people and females tend to have lower skills, employed people higher, and skills inversely correlate with geographic and linguistic distance to immigrants' birthplaces or native tongues.

Table 3.5 shows regression coefficients for a basic regression of language skills on the demographic characteristics described in Table 3.2. Six different measures of fluency are examined: binary measures of speaking and writing; the average of speaking, reading, writing, and comprehending; the baseline latent factor before DIF purification; and the latent factor purified on tenure within birthplace groups and purified on birthplace within tenure groups. The Table agrees with past studies on the signs of most variables, including the concavity in tenure and the interaction between tenure and age at migration. Age at migration has a positive coefficient, which is evidence of strong cohort effects among immigrants to France; the positive sign disappears if the full sample is used, rather than only immigrants who arrived after age 18.

Comparing coefficients across different measures of language skills, it is evident that different measures of skill yield different inferences. Comparing the probit coefficients, regressors correlate differently with speaking skills than than with writing skills; for instance, an extra of tenure corresponds to a greater increase in speaking fluency than in writing fluency. The OLS coefficients are more similar across regression specifications; the continuous latent trait estimates do not yield coefficients that are much different from those based on a simple average of reading, writing, speaking, and comprehending. A t-test on the equality of coefficients does not find any statistically significant differences between any of the tenure or age coefficients from any two OLS regressions in Table 3.5.[2]

2. The single exception is the tenure coefficient for the last column, which is statistically significantly different from the others.

Table 3.5: Determinants of Language Fluency

	Speak Well	Write Well	Average Self-Response (1-4 scale)	Raw L_n Estimate	Tenure purified by Birthplace	Birthplace purified by Tenure
	(probit)	(probit)	(OLS)	(OLS)	(OLS)	(OLS)
Age at Migration (AaM)	0.0168*	0.00689	0.0307*	0.0288*	0.0250	0.0255
	(0.00784)	(0.00753)	(0.0127)	(0.0131)	(0.0141)	(0.0135)
$(\text{AaM})^2/100$	-0.0291**	-0.0121	-0.0524**	-0.0529**	-0.0495**	-0.0480**
	(0.0109)	(0.0103)	(0.0174)	(0.0178)	(0.0191)	(0.0183)
Yrs Since Migration (YSM)	0.0342***	0.0239***	0.0498***	0.0522***	0.0509***	0.0276**
	(0.00569)	(0.00558)	(0.00957)	(0.00972)	(0.0105)	(0.0103)
$(\text{YSM})^2/100$	-0.0345***	-0.0261***	-0.0391**	-0.0464***	-0.0405**	0.0144
	(0.00729)	(0.00709)	(0.0120)	(0.0120)	(0.0130)	(0.0129)
(AaM)×(YSM)	-0.000502***	-0.000264	-0.000933***	-0.000843**	-0.000732*	-0.000814**
	(0.000152)	(0.000149)	(0.000260)	(0.000266)	(0.000287)	(0.000280)
Employed	0.105***	0.103***	0.248***	0.260***	0.265***	0.259***
	(0.0134)	(0.0131)	(0.0226)	(0.0231)	(0.0250)	(0.0244)
Married	-0.0110	-0.0237	-0.0519*	-0.0436	-0.0433	-0.0555*
	(0.0166)	(0.0157)	(0.0246)	(0.0254)	(0.0284)	(0.0275)
Endogamous	-0.150***	-0.110***	-0.263***	-0.277***	-0.300***	-0.291***
	(0.0132)	(0.0130)	(0.0215)	(0.0221)	(0.0243)	(0.0237)
Female	0.00806	0.0269*	0.0280	0.0223	0.0341	0.0351
	(0.0131)	(0.0126)	(0.0206)	(0.0211)	(0.0229)	(0.0225)
H.S. Degree	0.245***	0.285***	0.564***	0.570***	0.636***	0.621***
	(0.0158)	(0.0145)	(0.0254)	(0.0268)	(0.0294)	(0.0290)
College Degree	0.387***	0.419***	0.759***	0.785***	0.880***	0.867***
	(0.0149)	(0.0128)	(0.0215)	(0.0227)	(0.0252)	(0.0244)
Has Children	-0.0187	-0.00357	-0.0466	-0.0462	-0.0451	-0.0457
	(0.0190)	(0.0180)	(0.0282)	(0.0290)	(0.0319)	(0.0312)
MENA	0.0881***	0.209***	0.0917***	0.109***	0.258***	0.116***
	(0.0166)	(0.0160)	(0.0257)	(0.0260)	(0.0284)	(0.0279)
Sub-Sahara	0.211***	0.327***	0.361***	0.410***	0.106**	0.368***
	(0.0189)	(0.0171)	(0.0283)	(0.0286)	(0.0330)	(0.0310)
Other	-0.158***	-0.0469*	-0.222***	-0.264***	0.0637	-0.286***
	(0.0195)	(0.0197)	(0.0292)	(0.0299)	(0.0343)	(0.0328)
High Immigrant Density	0.0378	0.0875***	0.0341	0.0543	0.0754	0.0594
	(0.0259)	(0.0252)	(0.0409)	(0.0422)	(0.0459)	(0.0443)
(High Density)×(YSM)	-0.00421***	-0.00506***	-0.00484*	-0.00543**	-0.00684**	-0.00611**
	(0.00127)	(0.00121)	(0.00196)	(0.00200)	(0.00219)	(0.00213)
Constant			1.213***	-1.044***	-1.123***	-0.876***
			(0.219)	(0.227)	(0.244)	(0.235)
N	5031	5031	4987	4889	4889	4889
R^2			0.306	0.326	0.295	0.315

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

To compare OLS and probit coefficients, we must commensurate marginal outcomes from a probit regression with OLS coefficients. Consider the following back-of-the-envelope calculation. One extra year of tenure corresponds to a roughly 3.4% higher probability of speaking well. If this transition in skill levels were a Poisson process, we would expect roughly 30 years before the average person learned to speak well. For the continuous measure, the linear coefficient on L is approximately 0.05, and the quadratic term has a coefficient of -0.04. Over 30 years, the difference in L would be 1.14. Based on Figure 3.3, the 1.14 is smaller than the difference in L between the average poor speaker and average good speaker, smaller than the difference between the average respondents in Regions A and C, and larger than the difference between the average respondents in Regions A and B. Thus, the probit regression seems to mis-estimate the rate of increase in fluency, for various reasons depending on the individual – the basic point being that differences in average respondents do not accurately capture the distinctions between those who are fully fluent and those who speak well but do not have other linguistic skills.

Neighborhood effects are also a concern in studying language skills. Particularly for policy purposes, it is important to understand whether the density of fellow immigrants in one's neighborhood correlates with the level or rate of change of language skills. Past studies have found that a higher density of immigrants corresponds to a lower level of skill and a lower rate of learning the language. According to Table 3.5, immigrants in neighborhoods with a high density of other immigrants have higher fluency, on average, but acquire it more slowly over time. These patterns are detailed more closely in Table 3.6, where regressions are run separately for immigrants from different regions of the world. The table compares marginal effects for probit regressions on binary speaking proficiency and OLS coefficients on the baseline latent trait. Among Middle Eastern and African immigrants, those who live among a high density of people from the same part of the world have lower levels of language skill, with little effect on the rate of learning; the reverse is true for those who live amid high densities from other parts of the world.

Table 3.6: Effects of Local Immigrant Density on Fluency

	MENA Region		Sub-Sahara		Europe	
	Speak Well	$\hat{L}_n$	Speak Well	$\hat{L}_n$	Speak Well	$\hat{L}_n$
Age at Migration (AaM)	-0.00401*	-0.0134***	-0.00179	-0.00202	-0.0140***	-0.0216***
	(0.00184)	(0.00367)	(0.00237)	(0.00370)	(0.00191)	(0.00316)
Yrs Since Migration (YSM)	0.00637*	0.00874	-0.00424	-0.00915	0.00853**	0.0146***
	(0.00274)	(0.00504)	(0.00449)	(0.00688)	(0.00296)	(0.00436)
High Maghrebi Density	-0.0333	-0.102	0.0258	0.0870	-0.0728	-0.131
	(0.0478)	(0.0890)	(0.0577)	(0.0893)	(0.0546)	(0.0847)
(High Maghrebi)×(YSM)	-0.00272	-0.00282	-0.00193	-0.00669	-0.00156	-0.00166
	(0.00242)	(0.00427)	(0.00343)	(0.00535)	(0.00267)	(0.00385)
High Sub-Saharan Density	0.129**	0.368***	-0.154*	-0.221*	0.0970	0.201*
	(0.0432)	(0.0801)	(0.0665)	(0.0952)	(0.0566)	(0.0868)
(High Sub-Sahara)×(YSM)	-0.0000962	-0.00477	0.00303	0.00663	-0.00420	-0.00619
	(0.00223)	(0.00391)	(0.00390)	(0.00583)	(0.00276)	(0.00389)
High EU Density	0.0839	0.144	0.0353	0.0449	0.132*	0.240**
	(0.0451)	(0.0844)	(0.0591)	(0.0867)	(0.0520)	(0.0825)
(High Europe)×(YSM)	-0.00228	-0.00318	0.00364	0.00719	-0.00928***	-0.0159***
	(0.00228)	(0.00409)	(0.00355)	(0.00532)	(0.00279)	(0.00404)
Female	-0.121***	-0.265***	-0.0840**	-0.212***	0.103***	0.256***
	(0.0223)	(0.0420)	(0.0278)	(0.0449)	(0.0280)	(0.0434)
Constant		0.145		0.637***		0.200
		(0.150)		(0.156)		(0.152)
N	1840	1794	1093	1065	1198	1169
R^2		0.0655		0.0476		0.124

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

The value of language skills, rather than the determinants, is most important from an economic point of view. Table 3.7 shows coefficients for a regression of log-hourly wages on demographic characteristics and different measures of fluency, only for working men. The signs are the same as those found in previous studies: wages are positively correlated with migrating early in life, with longer tenures, with being European, and with education.

However, the wage premium for language skills depends on how those skills are measured. The binary measures of skill imply a 7.4% premium for speaking or writing well. For continuous measures, a one-unit increase in L results in approximately 5% higher wage. Thus, scaling to account for the difference between average poor- and high-skilled speakers of 1.45, this is about a 7.5% premium – matching the implied difference based on the binary regressor. But comparing an average poor speaker with someone in Region B, the implied wage premium is barely 5%; comparing an average poor speaker with someone in Region C, the implied wage difference is closer to 9%. In other words, those who speak well make approximately 5% more than those who do not, while those who speak, write, read, and comprehend earn an additional 4%.

These numbers imply that the average wage premium is mis-measured: it overestimates the returns to speaking skills alone, while obscuring the additional premium for writing, reading, and comprehending skills. These premia are not insubstantial: the additional premium for writing, reading, and comprehending is almost as large as the premium just for speaking.

Table 3.7: Determinants of log-Wages for Working Men

	Measure of Fluency					
	Speak Well	Write Well	Average Self-Response (1-4 scale)	Raw L_n Estimate	Tenure purified by Birthplace	Birthplace purified by Tenure
Fluency	0.0741*	0.0735*	0.0559*	0.0477*	0.0440*	0.0493*
	(0.0319)	(0.0340)	(0.0219)	(0.0206)	(0.0200)	(0.0202)
Age at Migration (AaM)	-0.0277	-0.0271	-0.0274	-0.0238	-0.0240	-0.0239
	(0.0158)	(0.0157)	(0.0158)	(0.0158)	(0.0158)	(0.0158)
$(\text{AaM})^2/100$	0.0539*	0.0528*	0.0536*	0.0466	0.0470	0.0468
	(0.0252)	(0.0251)	(0.0253)	(0.0251)	(0.0251)	(0.0251)
Yrs Since Migration (YSM)	0.0135*	0.0137*	0.0132*	0.0134*	0.0136*	0.0146*
	(0.00617)	(0.00617)	(0.00625)	(0.00631)	(0.00633)	(0.00628)
$(\text{YSM})^2/100$	-0.0179	-0.0182	-0.0173	-0.0179	-0.0184	-0.0208
	(0.0147)	(0.0147)	(0.0148)	(0.0150)	(0.0150)	(0.0150)
MENA	-0.161***	-0.172***	-0.164***	-0.169***	-0.175***	-0.169***
	(0.0326)	(0.0338)	(0.0327)	(0.0331)	(0.0337)	(0.0332)
Sub-Sahara	-0.235***	-0.249***	-0.243***	-0.252***	-0.237***	-0.250***
	(0.0339)	(0.0360)	(0.0340)	(0.0347)	(0.0336)	(0.0344)
Other	-0.240***	-0.253***	-0.245***	-0.254***	-0.270***	-0.252***
	(0.0628)	(0.0644)	(0.0649)	(0.0664)	(0.0671)	(0.0662)
H.S. Degree	0.104**	0.101**	0.0977**	0.102**	0.101**	0.0984**
	(0.0333)	(0.0342)	(0.0365)	(0.0363)	(0.0368)	(0.0367)
College Degree	0.357***	0.353***	0.347***	0.353***	0.352***	0.347***
	(0.0462)	(0.0490)	(0.0495)	(0.0496)	(0.0511)	(0.0510)
Constant	2.480***	2.491***	2.401***	2.484***	2.491***	2.481***
	(0.234)	(0.235)	(0.231)	(0.241)	(0.243)	(0.241)
N	1544	1544	1532	1503	1503	1503
R^2	0.107	0.106	0.108	0.107	0.107	0.108

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

3.6 Conclusion

The main lessons of this study are twofold: to improve the measurement of language skills and analysis of their economic effects, researchers must incorporate multiple dimensions of language skills. At the very least, they must measure skills using evidence of both verbal and non-verbal aspects of linguistic fluency: speaking as well as reading, writing, or understanding. But to do that, surveys must capture these multiple dimensions. So far, many large-scale surveys ask only about speaking, if they ask anything at all.

There is good news: based on the evidence presented here, there is not an especially high bar for attaining a measure of fluency that captures multiple dimensions of language skills and is robust to DIF – a simple average of reading, writing, speaking, and comprehending (all measured on the same ordinal scale) yields the same regression results as a more sophisticated latent factor calculated via GRM. More work could explore whether the evidence from other surveys is similar, but this provides a benchmark for future studies.

CHAPTER 4

CONCLUSION

In this dissertation I have attempted to shed light on what I view as problematic open questions within both the narrow subfield of human capital and the broader field of economics as a whole. Answering these questions will depend on the availability of empirically-relevant theory as well as useful ways to measure social phenomena in order to test those theories. I have offered piecemeal contributions to both of these, with the empirical tools being guided by the predictions in my theory.

Human capital theory has made great strides in understanding various investments (broadly understood) through a cost-benefit framework, including schooling, on-the-job training, health, marriage, and child-bearing. The same cost-benefit framework has been applied to the important question of individuals' decision to migrate within or between countries, and to the subsequent decision of whether to acquire specific traits that will increase their wage or utility in their new home. In the most basic sense, then, this dissertation builds directly on past models by adding a few bells and whistles – a continuous time setting, multiple human capital stocks – and deriving a much richer set of possible outcomes, including two qualitatively distinct patterns of investment. The predictions point to a need for more detailed empirical analysis of immigrants' traits, and I apply standard tools from other fields to enrich the measurement of social outcomes, offering support for the theory along the way.

Yet I view this project as more than a contribution to immigration economics and human capital. Economics as a whole has been increasingly unable to rationalize (in both an economic and a mainstream sense) many of the driving forces of the modern world. It appears we are in an age of entrenched political polarization; an era in which people use the phrase "identity politics" to connote both positive and negative ideas; a world in which some of the most pressing political issues boil down to moral questions of rights. Who rightfully deserves a place in our society, who constitutes "Us" and "Them?" We can claim that there is no economic content to such questions, but on the contrary, these are fundamentally questions of taste. Our tastes determine not just what we consume but also the people with whom we associate, what we consider morally right versus wrong (or in the typical language of tastes: tolerable versus disgusting) and many of the behaviors, such as religious practices, in which we engage on a daily basis. And it is almost a truism in microeconomics that preferences and tastes are the fundamental building blocks of our human capital; in the most basic sense they define us as individuals.

Therefore, it seems to me that the very important social-scientific phenomena of political polarization, group formation, and cultural change must have a human capital explanation. Models that treat relationships or identities as discrete choices that can be toggled on and off (even at some cost) are, at their core, models of taste as a discrete choice that can be here one day, gone tomorrow, whatever is convenient. Such models cannot account for the richness of responses to social forces such as those documented by Richerson & Boyd (2005).

Yet human capital theory, if it has failed anywhere, has failed to model the formation of preferences and tastes. I hope that the work is only beginning (and, hopefully, not ending with the seminal work of Stigler & Becker (1977) and Becker & Murphy (1988)). Much more work needs to be done, if human capital models really are to help us understand political and social history, and in turn to guide us in the future. I believe that my model, which is a reincarnation of Stigler and Becker's original model of taste formation, can be generalized to include social interactions and ultimately to be applied outside of immigration to explain cultural change, group formation, why culture hasn't caught up to globalization. What my model shows, at the very least, is that multiple types of outcomes can result from the same process, depending on the relevant elasticities. It offers a blueprint for thinking about social interactions along the same lines, and possibly for extending traditional social network models to incorporate human capital investment decisions. Such extensions would make existing models more realistic while also explaining a wider variety of phenomena via a single theoretical framework.

APPENDIX A

PROOFS OF RESULTS

1 Result 1

To determine stability of a steady-state, examine the determinant of the Jacobian. For completeness, we derive the Jacobian in its entirety, first by differentiating equation 1.7 and using the fact that $\ell = 1 - m$ and $L = \frac{1}{\delta} - M$. The derivatives are:

$$\frac{d\dot{\ell}}{d\ell} = \frac{\left(u_{\ell\ell L} - u_{mmM}\right)\left(\ell - \delta L\right) + u_{\ell L} + u_{mM} - u_{\ell L} - u_{mM} - \left(\rho + \delta\right)\left(u_{\ell\ell} + u_{mm}\right)}{-u_{\ell\ell} - u_{mm}} + \dot{\ell}\frac{u_{mmm} - u_{\ell\ell\ell}}{\left(u_{\ell\ell} + u_{mm}\right)^2}$$

$$\frac{d\dot{\ell}}{dL} = \frac{\left(u_{\ell LL} - u_{mMM}\right)\left(\ell - \delta L\right) - \delta\left(u_{\ell L} + u_{mM}\right) - u_{LL} - u_{MM} - \left(\rho + \delta\right)\left(u_{\ell L} + u_{mM}\right)}{-u_{\ell\ell} - u_{mm}} + \dot{\ell}\frac{u_{mmM} - u_{\ell\ell L}}{\left(u_{\ell\ell} + u_{mm}\right)^2}$$

$$\frac{d\dot{L}}{d\ell} = 1$$

$$\frac{d\dot{L}}{dL} = -\delta$$

Imposing the steady-state conditions and simplifying reduces these formulas allows us to write the linear approximation to the steady state as in equation 1.1.

$$\begin{pmatrix} \Delta\dot{\ell} \\ \Delta\dot{L} \end{pmatrix} = \begin{pmatrix} \rho+\delta & \frac{(u_{LL}+u_{MM})+(\rho+2\delta)(u_{\ell L}+u_{mM})}{u_{\ell\ell}+u_{mm}} \\ 1 & -\delta \end{pmatrix} \begin{pmatrix} \ell-\ell^* \\ L-L^* \end{pmatrix} \tag{1.1a}$$

$$= \begin{pmatrix} \rho+\delta & J \\ 1 & -\delta \end{pmatrix} \begin{pmatrix} \ell-\ell^* \\ L-L^* \end{pmatrix} \tag{1.1b}$$

Let J be the upper-right element of the Jacobian. The trace of the steady-state Jacobian is positive, so the system has two unstable (positive) roots if and only if the determinant is also positive. This happens if and only if $-\delta\,(\rho+\delta) > J$. Rearranging this formula yields point (*1*). It follows that $J < 0$ – equivalent to adjacent complementarity – is necessary but not sufficient, proving point (*2*). □

2 Result 2

According to differential equations 1.7, a steady state must satisfy $\ell = \delta L$ and $u_M - u_L = (\rho+\delta)(u_\ell - u_m)$. Clearly, if $u_M = u_L$ and $u_\ell = u_m$ then the latter condition will hold. Start by solving for $0 = u_L - u_M$ with the condition that $\ell = \delta L$:

$$\begin{aligned} 0 &= \alpha\beta\,(\delta L)^\gamma L^{\beta-1} - (1-\alpha)\,\beta\,(1-\delta L)^\gamma \left(\frac{1}{\delta} - L\right)^{\beta-1} \\ &= \alpha\,(\delta L)^{\gamma+\beta-1} - (1-\alpha)\,(1-\delta L)^{\gamma+\beta-1} \end{aligned}$$

The resulting steady state formulae are shown in equations 2.1. Substituting this values directly shows that they also set $u_\ell = u_m$, so the steady-state conditions are indeed satisfied.

$$L^* = \frac{1}{\delta}\frac{x}{x+1} \tag{2.1a}$$

$$\ell^* = \frac{x}{1+x} \tag{2.1b}$$

$$x = \left(\frac{1-\alpha}{\alpha}\right)^{\frac{1}{\gamma+\beta-1}} \tag{2.1c}$$

To verify uniqueness, note that the steady-state conditions can alternatively be satisfied if $-\frac{u_L - u_M}{u_\ell - u_m} = \rho + \delta$, i.e. the MRS in the constrained state space equals $\rho + \delta$. Calculating the MRS directly along the locus $\ell = \delta L$, we find:

$$\begin{aligned}
-\frac{u_L - u_M}{u_\ell - u_m}\bigg|_{\ell=\delta L} &= -\frac{\alpha\beta\left(\delta L\right)^{\gamma} L^{\beta-1} - (1-\alpha)\,\beta\left(1-\delta L\right)^{\gamma}\left(\frac{1}{\delta} - L\right)^{\beta-1}}{\alpha\gamma\left(\delta L\right)^{\gamma-1} L^{\beta} - (1-\alpha)\,\gamma\left(1-\delta L\right)^{\gamma-1}\left(\frac{1}{\delta} - L\right)^{\beta}} \\
&= -\frac{\beta}{\gamma}\cdot\frac{\alpha\left(\delta L\right)^{\gamma+\beta-1} - (1-\alpha)\left(1-\delta L\right)^{\gamma+\beta-1}}{\alpha\left(\delta L\right)^{\gamma+\beta-1} - (1-\alpha)\left(1-\delta L\right)^{\gamma+\beta-1}} \\
&= -\frac{\beta}{\gamma}
\end{aligned}$$

Thus, for $\beta, \gamma > 0$ this can never equal $\rho + \delta$. The steady state in equation 2.1 is therefore unique, proving (*1*).

Next, note that the conditions $u_L = u_M$ and $u_\ell = u_m$ imply that the gradient of the constrained utility function must equal zero. Hence, the steady state is a local extremum or a saddle point, proving (*2*). Finally, calculate the determinant of the Hessian at the steady state:

$$\begin{aligned}
D &= \left(u_{LL} + u_{MM}\right)\left(u_{\ell\ell} + u_{mm}\right) - \left(u_{\ell L} + u_{mM}\right)^2\Big|_{\ell^*, L^*} \\
&= \left[(\gamma-1)(\beta-1) - \gamma\beta\right]\gamma\beta\frac{1}{\delta^{2\beta-2}}\left[\alpha\left(\frac{x}{1+x}\right)^{\gamma+\beta-2} + (1-\alpha)\left(\frac{1}{1+x}\right)^{\gamma+\beta-2}\right]^2
\end{aligned}$$

A saddle point occurs if $D < 0$, which happens if and only if $(\gamma-1)(\beta-1) < \gamma\beta$, which is equivalent to $\gamma+\beta>1$. In this case, utility increases towards the corners $(\ell, L)=(0,0)$ and $\left(1, \frac{1}{\delta}\right)$, implying a monolingual system. When $D>0$, then because $u_{\ell\ell}+u_{mm}<0$ we must have a local maximum. Utility is highest at the steady state, implying a bilingual system. This proves *(3)*.

Now consider the conditions determining stability. We can rewrite these conditions in terms of the roots of the Jacobian:

$$\mu=\frac{\rho}{2} \pm \frac{\sqrt{(\rho+2\delta)^{2}+4J}}{2} \tag{2.2}$$

$$=\frac{\rho}{2} \pm \frac{\sqrt{(\rho+2\delta)^{2}+4(\rho+2\delta) \frac{u_{\ell L}+u_{mM}}{u_{\ell\ell}+u_{mm}}+4 \frac{u_{LL}+u_{MM}}{u_{\ell\ell}+u_{mm}}}}{2} \tag{2.3}$$

$$=\frac{\rho}{2} \pm \frac{\sqrt{X}}{2} \tag{2.4}$$

The formula J is the upper-right term in the Jacobian, as in equation 1.1. For both roots to be unstable, the discriminant X in the quadratic formula must be less than ρ^{2}. Since X is itself a quadratic form in $\rho+2\delta$, a necessary condition for being less than ρ^{2} is as follows:

$$X<\rho^{2} \Leftrightarrow \rho+2\delta<-2 \frac{u_{\ell L}+u_{mM}}{u_{\ell\ell}+u_{mm}}+\frac{1}{2} \sqrt{16\left(\frac{u_{\ell L}+u_{mM}}{u_{\ell\ell}+u_{mm}}\right)^{2}-16 \frac{u_{LL}+u_{MM}}{u_{\ell\ell}+u_{mm}}+4\rho^{2}}$$

$$\begin{aligned}=&-2 \frac{u_{\ell L}+u_{mM}}{u_{\ell\ell}+u_{mm}} \\ &+\left|\frac{2}{u_{\ell\ell}+u_{mm}}\right| \sqrt{\left(u_{\ell L}+u_{mM}\right)^{2}-\left(u_{LL}+u_{MM}\right)\left(u_{\ell\ell}+u_{mm}\right)+\frac{\rho^{2}}{4}\left(u_{\ell\ell}+u_{mm}\right)^{2}}\end{aligned}$$

$$=-2 \frac{u_{\ell L}+u_{mM}}{u_{\ell\ell}+u_{mm}}+\left|\frac{2}{u_{\ell\ell}+u_{mm}}\right| \sqrt{\frac{\rho^{2}}{4}\left(u_{\ell\ell}+u_{mm}\right)^{2}-D}$$

Simplify the condition by substituting the steady-state formulae for this particular utility function:

$$X = \rho^2 \Leftrightarrow \rho + 2\delta < -\frac{2\gamma\beta\delta}{\gamma(\gamma-1)}$$
$$+ \frac{2}{\gamma(1-\gamma)\left(\frac{1}{\delta}\right)^{\beta}} \sqrt{\frac{\rho^2}{4}\gamma^2(\gamma-1)^2\left(\frac{1}{\delta}\right)^{2\beta} - \left[(\gamma-1)(\beta-1) - \gamma\beta\right]\gamma\beta\left(\frac{1}{\delta}\right)^{2\beta-2}}$$

$$= \frac{2\beta\delta}{1-\gamma} + \frac{2\beta\delta}{1-\gamma}\sqrt{\frac{\rho^2(1-\gamma)^2}{4\beta^2\delta^2} - \frac{(\gamma-1)(\beta-1)-\gamma\beta}{\gamma\beta}}$$

At the borderline case where $D = 0$ then $\gamma + \beta = 1$ and the formula simplifies to $\rho + 2\delta < 2\delta\frac{\beta}{1-\gamma} + \rho$, which is not strictly satisfied but rather holds with equality. This is therefore a knife-edge case. If $D > 0$ then the square root will be smaller than when $D = 0$, so the second term will be less than ρ (or possibly complex if the square root has a negative argument). Further, $\beta + \gamma < 1$, so $\frac{\beta}{1-\gamma} < 1$ and the first term will be less than 2δ. Hence, the inequality can never hold. Finally, if $D < 0$ then the second term will be larger than ρ. Further, $\beta + \gamma > 1$, so $\frac{\beta}{1-\gamma} > 1$ and the first term will be larger than 2δ. Thus, the inequality is always satisfied.

For the case where $D < 0$, we must now check the other necessary condition, i.e. $\rho + 2\delta$ must be larger than the other root of the quadratic formula:

$$\rho + 2\delta > \frac{2\beta\delta}{1-\gamma} - \frac{2\beta\delta}{1-\gamma}\sqrt{\frac{\rho^2(1-\gamma)^2}{4\beta^2\delta^2} - \frac{(\gamma-1)(\beta-1)-\gamma\beta}{\gamma\beta}}$$

Break up the square root:

$$\frac{2\beta\delta}{1-\gamma}\sqrt{\frac{\rho^2(1-\gamma)^2}{4\beta^2\delta^2} + \frac{\gamma\beta-(\gamma-1)(\beta-1)}{\gamma\beta}}$$
$$< \frac{2\beta\delta}{1-\gamma}\sqrt{\frac{\rho^2(1-\gamma)^2}{4\beta^2\delta^2}} + \frac{2\beta\delta}{1-\gamma}\sqrt{\frac{\gamma\beta-(\gamma-1)(\beta-1)}{\gamma\beta}}$$

$$< \rho + \frac{2\beta\delta}{1-\gamma}$$

$\square$

APPENDIX B

OTHER STEADY STATE CONFIGURATIONS

1 Multiple Interior Steady States

When there are multiple interior steady states, local dynamics are determined by stability of the nearby states. With multiple steady states, at least one will be stable and one will be unstable. The stable steady-states can be thought of as enclaves in which one language is dominant but the other is partially maintained, and many trajectories will lead toward these stable points. But the unstable steady-states will be generating most of the important dynamics, as trajectories flow away from these points; now, the stable steady states will take the place of true corner solutions. Therefore, in choosing the optimal trajectory a person is choosing between two different ratios of L and M, rather than pure monolingualism in one or the other. Thus, by and large, the addiction mechanism is still most salient when there are multiple steady states, and Regions 2 and 3 of Figure 1.4 become intertwined as trajectories run from unstable steady states towards stable ones.

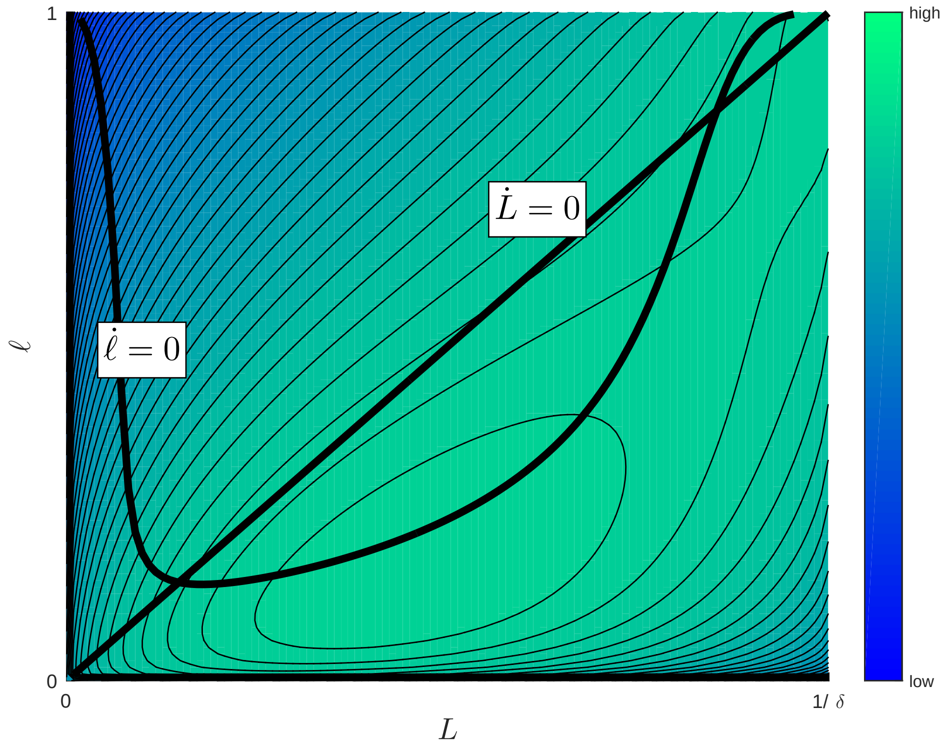
Figure B.1 illustrates a system with two steady states in which all three regions of Figure 1.4 are represented.[1] The lower steady state at L_1^* is stable, and represents an enclave in which language $\mathcal{M}$ is dominant, but some degree of langauge $\mathcal{L}$ is desirable. To the left (and for a small region to the right), distant complementarity holds (Region 1 of Figure 1.4). Between steady states adjacent complementarity holds. Based on the indifference curves (Panel (a)), the right steady state yields higher utility. But because it is unstable we would expect most trajectories will follow the stable arm towards the lower steady state, and the process would formally be characterized as bilingualism with addiction (Region 2 of Figure 1.4). But sufficiently far to the right, near $L = \frac{1}{\delta}$, trajectories may diverge, with some trajectories leading towards monolingualism in $\mathcal{L}$ (Region 3 of Figure 1.4). But again, the core mechanism in most of the state space is still addiction.

1. The utility function is $\ell^{0.1}L^{0.4} + 0.7m^{0.8}M^{0.7}$ with $\rho = 1$ and $\delta = 0.9$.

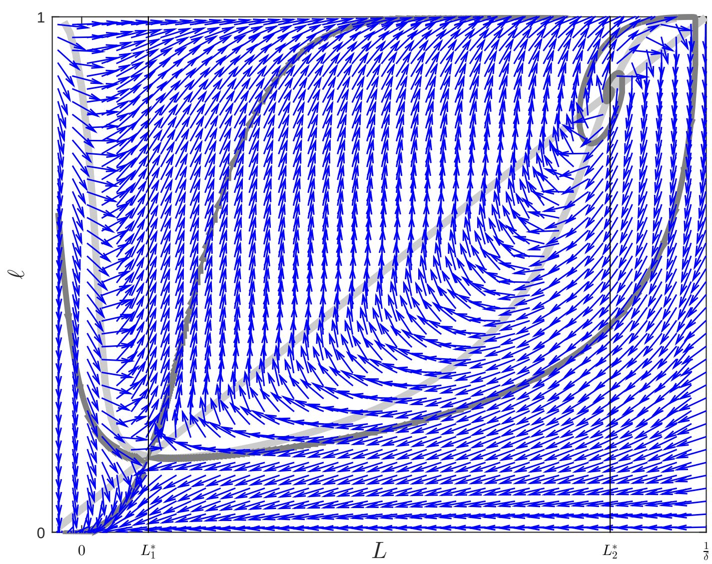
This graph is a combination of the three possible cases described in Section 1.3.2. More complicated cases could be concocted with more than two steady states, although the main intuition of Region 3 is still the important factor here.

Figure B.1: A SYSTEM WITH TWO INTERIOR STEADY STATES.

(a) INDIFFERENCE CURVES



(b) PHASE DIAGRAM



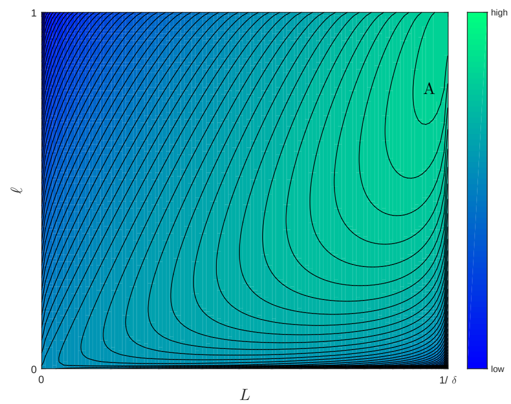
Indifference curves and phase diagram for a system with two interior steady states. All three types of dynamics from Figure 1.4 can appear here: distant complementarity (to the left of L_1^*), adjacent complementarity (to the right of L_1^*), or sufficiently strong that trajectories diverge (far right of L_2^*). Dark gray lines are invariant manifolds; light gray lines show the loci $\dot{\ell} = 0$ and $\dot{L} = 0$. The utility function is $\ell^{0.1}L^{0.4} + 0.7m^{0.8}M^{0.7}$ with $\rho = 1$ and $\delta = 0.9$.

2 No Interior Steady State

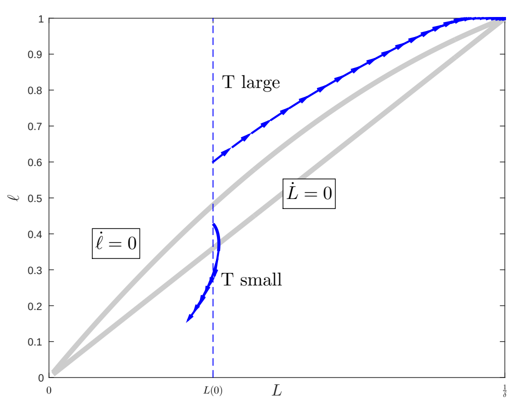
When there is no interior steady state, there cannot be a stable manifold leading to a long-run interior solution. Hence, every trajectory must lead toward a corner, yielding monolingualism and will appear qualitatively identical to Region 3 of Figure 1.4. Mathematically, there can be several manifestations of this. Figure B.2 offers one example, in which there is a global maximum (at point A) which is, however, unattainable as a steady state; all trajectories will lead towards a corner.

Figure B.2: A System with No Interior Steady State.

(a) Indifference Curves



(b) Trajectories



Indifference curves and sample trajectories for a system with a global maximum (point A) but no interior steady state. The utility function is $u = \ell^{\frac{2}{5}} L^{\frac{3}{5}} + 0.7 m^{\frac{4}{5}} M^{\frac{1}{5}}$.

3 Concavity with Unstable Steady State

When utility is strictly concave, the steady state can still be unstable; this means the Hessian and the Jacobian at the steady state have opposite signs. The instability must come from ρ and δ, which do not affect the Hessian but do determine intertemporal dynamics. When depreciation is sufficiently slow and people are sufficiently impatient (small δ and large ρ), then culture can appear sufficiently addictive that people prefer complete specialization in one stock or the other. The Cobb-Douglas utility function used previously will not work as an illustration, but quadratic utility will suffice.[2]

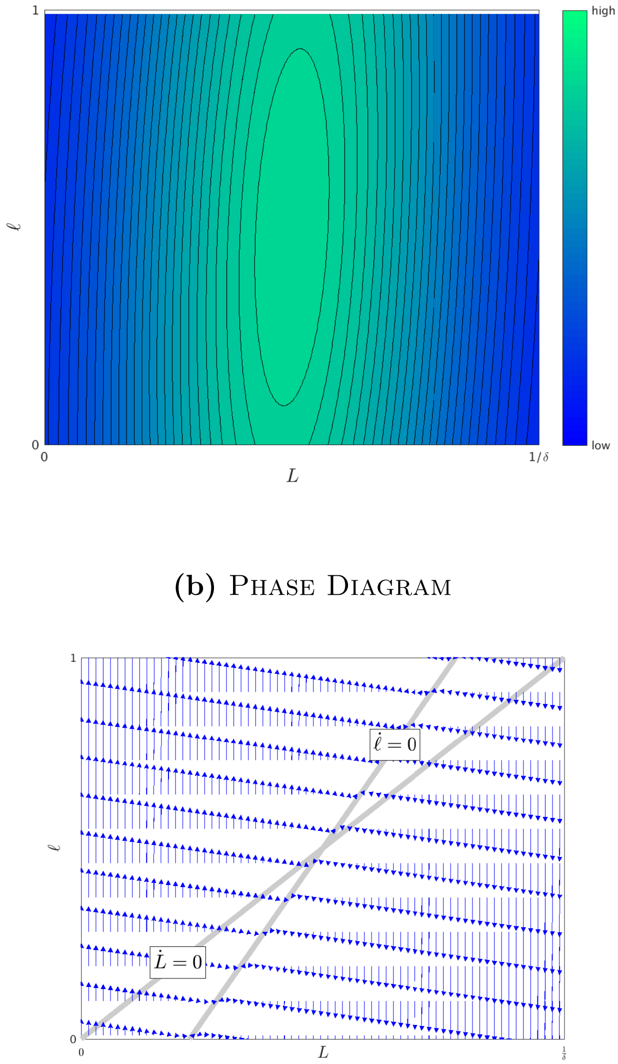
Consider, for example, Figure B.3. The top panel shows the indifference curves, verifying that this is indeed strictly concave with a global maximum at the midpoint of the feasible space. The lower panel shows the phase diagram; the steady state is evidently unstable, but the small arrows near the steady state show that a small deviation from this point will result in very slow divergence. In other words, it will appear to be locally stable for long periods of time. Far away from the steady state, however, trajectories do not lead anywhere near the steady state but instead lead towards an edge. Intuitively, stocks depreciate too slowly, and people are too impatient, for dynamics to lead "up the hill" toward the steady state; instead people coast along the indifference curves toward an edge where they exclusively speak one language until they have built complete fluency in it.

2. Quadratic utility can be considered as a second-order approximation to a more general utility function. For the example in this section, the utility function is

Figure B.3: A Concave System with Unstable Steady State.

(a) Indifference Curves

(b) Phase Diagram



Indifference curves and phase diagram for a system with strictly concave utility but an unstable steady state. The utility function is quadratic: $u =$.

APPENDIX C

EMPIRICAL APPENDIX

1 Linguistic Distance

To calculate linguistic proximities, I adapt the method of Fearon (2003, section 8), which in turn relies on the phylogenetic language trees of Lewis *et al.* (2016). Phylogenetic trees are basically branching diagrams showing the evolutionary relationships between various languages; each branch is labeled with a category, and linguistically similar languages share more branches. Characteristics like mutual intelligibility cannot necessarily be inferred baed on phylogenetic proximity. Nevertheless, phylogenetic trees have yielded numerous indexes, including distance data, that are robustly correlated with macroeconomic and political indicators of development (Alesina *et al.* , 2003; Spolaore & Wacziarg, 2009).

Suppose the relevant local language j (English, Spanish, or French) has N_j branches on the phylogenetic tree, a person's mother tongue i has N_i branches, and the two languages share n_{ij} branches. Linguistic distance for individual i's mother tongue in country j is given by d_i in equation 1.1.

$$d_i = \left(\frac{n_i}{\max\left(N_i, N_j\right)} \right)^{\sigma} \tag{1.1}$$

The number of total possible branches data ranges from 1 (for language isolates like Basque) to 14 (for Maori, which does not share any branches with English, Spanish, or French). The convexity parameter $\sigma \in [0, 1]$ determines the the relative weight given to different branching points. Languages that share fewer branches should be relatively more "distant" than languages that share nearly all branches. In the regressions for this paper, $\sigma = 0.5$, although the results are not sensitive to any mid-range value of σ.

2 Location-Scale Ordered Choice Model

The general location-scale model of Williams (2009) works as follows: an underlying (continuous) variable Y^* and its observed counterpart Y take the form

$$Y_i^* = X_i'\beta + \sigma\left(Z_i'\phi\right)\epsilon_i \tag{2.1a}$$

$$\epsilon_i \sim N(0,1) \tag{2.1b}$$

$$Y_i = \begin{cases} 1 & Y_i^* \leq \mu_1 \\ 2 & \mu_1 < Y_i^* \leq \mu_2 \\ \vdots & \\ N & \mu_N < Y_i^* \end{cases} \tag{2.1c}$$

The mean and variance are determined by X_i and Z_i, which may (but do not have to) share the same covariates. The conditional log-variance is directly proportional to $Z_i'\phi$:

$$\operatorname{Var}\left[\epsilon_i \middle| Z_i\right] \propto \exp\left(Z_i'\phi\right)^2 \tag{2.2}$$

Therefore, the variance increases with variable z_{ij} whenever $\phi_j > 0$.

When Y is binary, the usual marginal effects of a logit or probit are altered. For variables that appear in both X_i and Z_i the marginal effects are a mixture of effects from the variance and from the mean. This can produce ambiguous interpretations if one only looks at the regression coefficients.

In particular, for this study an immigrant's tenure appears in both X_i and Z_i. The overall marginal effect, then, is given by equation 2.3, where f is either a logit or probit density (see also Greene & Hensher, 2010, §7.4).

$$\frac{\partial \Pr\left\{Y_i = 1 \middle| X_i, Z_i\right\}}{\partial Tenure_i} = f\left(\frac{\mu_1 - X_i'\beta}{\exp\left(Z_i'\phi\right)}\right)\frac{\beta_{Tenure} + \left(\mu_1 - X_i'\beta\right)\phi_{Tenure}}{\exp\left(Z_i'\phi\right)} \tag{2.3}$$

REFERENCES

Abramitzky, Ran, Boustan, Leah, & Eriksson, Katherine. 2014 (September). *Cultural Assimilation During the Age of Mass Migration*. Preliminary Working Paper.

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

Akerlof, George A., & Kranton, Rachel E. 2010. *Identity Economics: How Identities Shape Our Work, Wages, and Well-Being*. Princeton University Press, Princeton.

Alesina, Alberto, & Spolaore, Enrico. 1997. On the Number and Size of Nations. *Quarterly Journal of Economics*, **112**(4), 1027–1056.

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

Allison, Paul D. 1999. Comparing Logit and Probit Coefficients Across Groups. *Sociological Methods & Research*, **28**(2), 186–208.

Andujo, Estela. 1988. Ethnic Identity of Transethnically Adopted Hispanic Adolescents. *Social Work*, **33**(6), 531–535.

Atkin, David. 2016. The Caloric Cost of Culture: Evidence from Indian Migrants. *American Economic Review*, **106**(4), 1144–1181.

Beauchemin, Cris, & Simon, Patrick. *TeO Project Description*. `http://teo_english.site.ined.fr/fichier/s_rubrique/20309/teo.note.eng.en.pdf`.

Becker, Gary S., & Murphy, Kevin M. 1988. A Theory of Rational Addiction. *Journal of Political Economy*, **96**(4), 675–700.

Becker, Gary S., Grossman, Michael, & Murphy, Kevin M. 1994. An Empirical Analysis of Cigarette Addiction. *American Economic Review*, **84**(3), 396–418.

Belli, Robert F., Shay, William L., & Stafford, Frank P. 2001. Event History Calendars and Question List Surveys: A Direct Comparison of Interviewing Methods. *Public Opinion Quarterly*, **65**, 45–74.

Ben-Porath, Yoram. 1967. The Production of Human Capital and the Life-Cycle of Earnings. *Journal of Political Economy*, **75**(4, Part 1), 352–365.

Berman, Eli, Lang, Kevin, & Siniver, Erez. 2003. Language-skill Complementarity: Returns to Immigrant Language Acquisition. *Labour Economics*, **10**(3), 265–290.

Bertrand, Marianne, & Mullainathan, Sendhil. 2004. Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *The American Economic Review*, **94**(4), 991–1013.

Bisin, Alberto, & Verdier, Thierry. 2000a. 'Beyond the Melting Pot:' Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits. *The Quarterly Journal of Economics*, **115**(3), 955–988.

Bisin, Alberto, & Verdier, Thierry. 2000b. A Model of Cultural Transmission, Voting and Political Ideology. *European Journal of Political Economy*, **16**(1), 5–29.

Bisin, Alberto, & Verdier, Thierry. 2001. The Economics of Cultural Transmission and the Dynamics of Preferences. *Journal of Economic Theory*, **97**(2), 298–319.

Bisin, Alberto, Topa, Giorgio, & Verdier, Thierry. 2004. Religious Intermarriage and Socialization in the United States. *Journal of Political Economy*, **112**(3), 615–664.

Bisin, Alberto, Patacchini, Eleonora, Verdier, Thierry, & Zenou, Yves. 2011. Formation and Persistence of Oppositional Identities. *European Economic Review*, **55**(8), 1046–1071.

Bisin, Alberto, Patacchini, Eleonora, Verdier, Thierry, & Zenou, Yves. 2016. Bend It Like Beckham: Ethnic Identity and Integration. *European Economic Review*, **forthcoming**.

Blau, Francine, Kahn, Lawrence, & Papps, Kerry. 2011. Gender, Source Country Characteristics and Labor Market Assimilation Among Immigrants: 1980-2000. *Review of Economics and Statistics*, **93**(1), 43–58.

Boada, Irene. 2015. How Catalan Survived. *The Atlantic*, September.

Borjas, George J. 1985. Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants. *Journal of Labor Economics*, **3**(4), 463–489.

Borjas, George J. 2000. The Economic Progress of Immigrants. *Chap. 1, pages 15–49 of:* Borjas, George J. (ed), *Issues in the Economics of Immigration*. University of Chicago Press.

Borjas, George J. 2014. *Immigration Economics*. Harvard University Press.

Borjas, George J. 2015. The Slowdown in the Economic Assimilation of Immigrants: Aging and Cohort Effects Revisited Again. *Journal of Human Capital*, **9**(4), 483–517.

Boyer, Marcel. 1983. Rational Demand and Expenditure Patterns under Habit Formation. *Journal of Economic Theory*, **31**(1), 27–53.

Castiglione, Concetta, & Infante, Davide. 2015. Rational Addiction and Cultural Goods: the Case of the Italian Theatregoer. *Journal of Cultural Economics*, April, 1–28.

Chiswick, Barry R. 1978. The Effect of Americanization on the Earnings of Foreign-Born Men. *Journal of Political Economy*, **86**(5), 897–921.

Chiswick, Barry R. 1991. Speaking, Reading, and Earnings Among Low Skilled Immigrants. *Journal of Labor Economics*, **9**(2), 149–170.

Chiswick, Barry R., & Miller, Paul W. 1995. The Endogeneity Between Language and Earnings: International Analyses. *Journal of Labor Economics*, **13**(2), 246–288.

Choi, Seung W., Gibbons, Laura E., & Crane, Paul K. 2011. lordif: An R Package for Detecting Differential Item Functioning Using Iterative Hybrid Ordinal Logistic Regression/Item Response Theory and Monte Carlo Simulations. *Journal of Statistical Software*, **39**(8), 1–30.

Crane, Paul K., Gibbons, Laura E., Jolley, Lance, & van Belle, Gerald. 2006. Differential Item Functioning Analysis With Ordinal Logistic Regression Techniques: DIFdetect and difwithpar. *Medical Care*, **44**(11), S115–S123.

Dockner, Engelbert J., & Feichtinger, Gustav. 1991. On the Optimality of Limit Cycles in Dynamic Economic Systems. *Journal of Economics*, **53**(1), 31–50.

Dockner, Engelbert J., & Feichtinger, Gustav. 1993. Cyclical Consumption Patterns and Rational Addiction. *The American Economic Review*, **83**(1), 256–263.

Dragone, Davide, Manaresi, Francesco, & Savorelli, Luca. forthcoming. Obesity and Smoking: Can We Catch Two Birds with One Tax? *Health Economics*.

Dustmann, Christian. 1994. Speaking Fluency, Writing Fluency and Earnings of Migrants. *Journal of Population Economics*, **7**(2), 133–156.

Dustmann, Christian. 1999. Temporary Migration, Human Capital, and Language Fluency of Migrants. *Scandinavian Journal of Economics*, **101**(2), 297–314.

Dustmann, Christian. 2003. Return Migration, Wage Differentials, and the Optimal Migration Duration. *European Economic Review*, **47**, 353–369.

Dustmann, Christian, & Fabbri, Francesca. 2003. Language Proficiency and Labour Market Performance of Immigrants in the UK. *The Economic Journal*, **113**(489), 695–717.

Dustmann, Christian, & Görlach, Joseph-Simon. 2016. The Economics of Temporary Migrations. *Journal of Economic Literature*, **54**(1), 98–136.

Dustmann, Christian, & van Soest, Arthur. 2001. Language Fluency and Earnings: Estimation with Misclassified Language Indicators. *The Review of Economics and Statistics*, **83**(4), 663–674.

Dustmann, Christian, & van Soest, Arthur. 2004. An Analysis of Speaking Fluency of Immigrants Using Ordered Response Models with Classification Errors. *Journal of Business and Economic Statistics*, **22**(3), 312–321.

Eurostat. 2015 (May). *Migration and Migrant Population Statistics*. `http://ec.europa.eu/eurostat/statistics-explained/index.php/Migration_and_migrant_population_statistics#Migrant_population`.

Fearon, James D. 2003. Ethnic and Cultural Diversity by Country. *Journal of Economic Growth*, **8**(2), 195–222.

Fernández, Raquel. 2013. Cultural Change as Learning: The Evolution of Female Labor Force Participation Over a Century. *The American Economic Review*, **103**(1), 472–500.

Froese, Paul. 2004. Forced Secularization in Soviet Russia: Why an Atheistic Monopoly Failed. *Journal for the Scientific Study of Religion*, **43**(1), 35–50.

Fryer, Roland G., & Levitt, Steven D. 2004. The Causes and Consequences of Distinctively Black Names. *The Quarterly Journal of Economics*, **119**(3), 767–805.

Gans, Herbert J. 1992. Comment: Ethnic Invention and Acculturation: A Bumpy-Line Approach. *Journal of American Ethnic History*, **11**(1), 45–52.

Glaeser, Edward L., Laibson, David, & Sacerdote, Bruce. 2002. An Economic Approach to Social Capital. *The Economic Journal*, **112**(483), F437–F458.

Gordon, Milton M. 1964. *Assimilation in American Life: The Role of Race, Religion, and National Origins*. Oxford University Press.

Greene, William H., & Hensher, David A. 2010. *Modeling Ordered Choices*. Cambridge University Press.

Grin, FranÇois. 1992. Towards a Threshold Theory of Minority Language Survival. *Kyklos*, **45**(1), 69–97.

Hausman, Jerry A., Abrevaya, Jason, & Scott-Morton, Fiona M. 1998. Misclassification of the Dependent Variable in a Discrete Response Setting. *Journal of Econometrics*, **87**(2), 239–269.

Heard, J. Norman. 1973. *White into Red: A Study of the Assimilation of White Persons Captured by Indians*. The Scarecrow Press, Inc.

Iannaccone, Laurence R. 1984. *Consumption Capital and Habit Formation with an Application to Religious Participation*. Ph.D. thesis, University of Chicago.

Iannaccone, Laurence R. 1986. Addiction and Satiation. *Economics Letters*, **21**(1), 95–99.

Jasso, Guillermina, Massey, Douglas S., Rosenzweig, Mark R., & Smith, James P. 2005. The U.S. New Immigrant Survey: Overview and Preliminary Results Based on the New-Immigrant Cohorts of 1996 and 2003. *In:* Morgan, Beverley, & Nicholson, Ben (eds), *Immigration Research and Statistics Service Workshop on Longitudinal Surveys and Cross-Cultural Survey Design: Workshop Proceedings*. Crown Publishing.

King, Gary, Murray, Christopher J. L., Salomon, Joshua A., & Tandon, Ajay. 2004. Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research. *American Political Science Review*, **98**(1), 191–207.

Kliewer, Erich. 1991. Immigrant Suicide in Australia, Canada, England and Wales, and the United States. *Journal of the Australian Population Association*, **8**(2), 111–128.

Kliewer, Erich V., & Ward, R. H. 1988. Convergence of Immigrant Suicide Rates to those in the Destination Country. *American Journal of Epidemiology*, **127**(3), 640–653.

Kónya, István. 2007. Optimal Immigration and Cultural Assimilation. *Journal of Labor Economics*, **25**(2), 367–291.

Lazear, Edward P. 1999. Culture and Language. *Journal of Political Economy*, **107**(6), S95–S126.

Lewis, M. Paul, Simons, Gary F., & Fennig, Charles D. 2016. *Ethnologue: Languages of the World, Nineteenth Edition*. Online Version http://www.ethnologue.com.

Lydens, Lois A. 1988. *A Longitudinal Study of Crosscultural Adoption: Identity Development Among Asian Adoptees at Adolescence and Early Adulthood*. PhD dissertation, Northwestern University.

Maher, Julianne. 1991. A Crosslinguistic Study of Language Contact and Language Attrition. *Chap. 5, pages 67–86 of:* Seliger, Herbert W., & Vago, Robert M. (eds), *First Language Attrition*. Cambridge University Press.

Marrone, James V. 2017. *Linguistic and Cultural Change as a Human Capital Process*. Working Paper.

McManus, Walter, Gould, William, & Welch, Finis. 1983. Earnings of Hispanic Men: The Role of English Language Proficiency. *Journal of Labor Economics*, **1**(2), 101–130.

McManus, Walter S. 1985. Labor Market Costs of Language Disparity: An Interpretation of Hispanic Earnings Differences. *The American Economic Review*, **75**(4), 818–827.

Mir, Conxita. 2008. The Francoist Repression in the Catalan Countries. *Catalan Historical Review*, **1**, 133–147.

Muraki, Eiji, & Carlson, James E. 1995. Full-Information Factor Analysis for Polytomous Item Responses. *Applied Psychological Measurement*, **19**(1), 73–90.

Pavlish, Carol Lynn, Noor, Sahra, & Brandt, Joan. 2010. Somali Immigrant Women and the American Health Care System: Discordant Beliefs, Divergent Expectations, and Silent Worries. *Social Science & Medicine*, **71**(2), 353–361.

Pew, Research Center. 2016 (July). *Europeans Fear Wave of Refugees Will Mean More Terrorism, Fewer Jobs*. Report.

Portes, Alejandro, & Zhou, Min. 1993. The New Second Generation: Segmented Assimilation and Its Variants. *The Annals of the American Academy of Political and Social Science*, **530**(1), 74–96.

Reckase, Mark D. 2009. *Multidimensional Item Response Theory*. New York: Springer.

Reher, David, & Requena, Miguel. 2009. The National Immigrant Survey of Spain: A new data source for migration studies The National Immigrant Survey of Spain: A New Data Source for Migration Studies in Europe. *Demographic Research*, **20**(March), 253–278.

Richerson, Peter J, & Boyd, Robert. 2005. *Not by Genes Alone: How Culture Transformed Human Evolution*. University of Chicago Press.

Rumbaut, Rubén G., & Portes, Alejandro. 2001. Ethnogenesis: Coming of Age in Immigrant America. *Chap. 1, pages 1–19 of:* Rumbaut, Rubén G., & Portes, Alejandro (eds), *Ethnicities: Children of Immigrants in America*. Russell Sage Foundation.

Ryder, Harl E. Jr., & Heal, Geoffrey M. 1973. Optimum Growth with Intertemporally Dependent Preferences. *Review of Economic Studies*, **40**(January), 1–33.

Samejima, Fumiko. 1969. *Estimation of Latent Ability Using a Response Pattern of Graded Scores*. Monograph Supplement, 34 Psychometric Monograph No. 17. Psychometric Society.

Schachtman, Tom. 2006. *Rumspringa: To Be or Not to Be Amish*. North Point Press.

Schmid, Monika S. 2002. *First Language Attrition, Use, and Maintenance: The Case of German Jews in Anglophone Countries*. John Benjamins.

Schmid, Monika S. 2011. *Language Attrition*. Key Topics in Sociolinguistics. Cambridge University Press.

Schmid, Monika S., Köpke, Barbara, Keijzer, Merel, & Weilemar, Lina (eds). 2004. *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues*. Studies in Bilingualism. John Benjamins.

Schoellman, Todd. 2010. The Occupations and Human Capital of U.S. Immigrants. *Journal of Human Capital*, **4**(1), 1–34.

Spolaore, Enrico, & Wacziarg, Romain. 2009. The Diffusion of Development. *The Quarterly Journal of Economics*, **124**(2), 469–529.

Spångberg, Emanuel Bylund. 2009. Maturational Constraints and First Language Attrition. *Language Learning*, **59**(3), 687–715.

Stigler, George J., & Becker, Gary S. 1977. De Gustibus Non Est Disputandem. *The American Economic Review*, **67**(2), 76–90.

Stolzenberg, Ross M. 1990. Ethnicity, Geography, and Occupational Achievement of Hispanic Men in the United States. *American Sociological Review*, **55**(1), 143–154.

Swaminathan, Hariharan, & Rogers, H. Jane. 1990. Detecting Differential Item Functioning Using Logistici Regression Procedures. *Journal of Educational Measurement*, **27**(4), 361–370.

Ventureyra, Valérie, & Pallier, Christophe. 2004. In Search of the Lost Language: The Case of Adopted Koreans in France. *Pages 207–221 of:* Schmid, Monika S., Köpke, Barbara, Keijzer, Merel, & Weilemar, Lina (eds), *Language Attrition: Interdisciplinary Perspectives on Methodological Issues*. John Benjamins.

Vigdor, Jacob. 2015. The Civic and Cultural Assimilation of Immigrants to the United States. *In:* Powell, Benjamin (ed), *The Economics of Immigration: Market-Based Approaches, Social Science, and Public Policy*. Oxford University Press. Unpublished.

White, Michael J., & Glick, Jennifer E. 2009. *Achieving Anew: How New Immigrants Do in American Schools, Jobs, and Neighborhoods*. Russell Sage Foundation.

Williams, Richard. 2009. Using Heterogeneous Choice Models to Compare Logit and Probit Coefficients Across Groups. *Sociological Methods & Research*, **37**(4), 531–559.

Zumbo, Bruno D. 2007. Three Generations of DIF Analyses: Considering Where It Has Been, Where It Is Now, and Where It Is Going. *Language Assessment Quarterly*, **4**(2), 223–233.