CSEF

Centre for Studies in Economics and Finance

WORKING PAPER NO. 799

Dynamic Ethnosizing and Revisable Identity

Nicodemo de Vito, Fausto Galli, Giuseppe Russo, Ružica Savčić, and Paolo Solimene

October 2026

University of Naples Federico II

University of Salerno

Bocconi University, Milan

CSEF - Centre for Studies in Economics and Finance
DEPARTMENT OF ECONOMICS AND STATISTICS – UNIVERSITY OF NAPLES FEDERICO II
80126 NAPLES - ITALY
Tel. and fax +39 081 675372 – e-mail: csef@unina.it
ISSN: 2240-9696

CSEF
Centre for Studies in Economics and Finance

WORKING PAPER NO. 799

Dynamic Ethnosizing and Revisable Identity

Nicodemo de Vito[*], Fausto Galli[†], Giuseppe Russo[‡], Ružica Savčić[§], and Paolo Solimene[**]

Abstract

Using panel data from the German SOEP (2008–2019), we transform the Ethnosizer (Constant et al., 2009) from a cross-sectional measure of immigrant identity into a dynamic tool that tracks identity shifts over time. Methodologically, we estimate a simultaneous two-equation dynamic binary-response model with correlated random effects, which allows us to recover state dependence and transition probabilities between the identity states defined by Berry (1980). Our results show that commitments to destination and origin cultures are highly persistent once formed —yet they are not irreversible. Transitions do occur and, among them, we observe a phenomenon of "integration reversal", whereby identity moves from previously integrated positions to more separated ones. This reversibility of the integration process demonstrates that integration is not a final destination, but a path-dependent process shaped by persistence, contextual shocks, and individual experiences. We complement the empirical analysis with a theoretical model that provides a foundation for the mechanisms underlying identity revision and integration reversal.

Keywords: Identity formation, immigration, acculturation, cultural transmission.

JEL classification: D91, F22, J15, K37, Z13.

Acknowledgments: We thank Pierpaolo Battigalli, Annamaria Menichini, Chris Muris, Aikaterini Kyriazidou and Peter Wakker for the valuable insights. We also thank participants to the GLO-Guangzhou 2026 Conference, the Petralia Workshop in Applied Economics, the CRETE 2026, the Final Spoke 3 Workshop in Naples, the 2026 EPCS Meeting in Madrid, and seminar participants at the University of Cyprus, the Xiamen University and the Tinbergen Institute.

[*] University of Salerno and Bocconi University.

[†] University of Salerno.

[‡] University of Salerno, CSEF and GLO. E-mail: grusso@unisa.it

[§] University of Cyprus.

[**] University of Rotterdam and Tinbergen Institute.

1 Introduction

The process of immigrant incorporation into host societies inevitably leads to some form of integration, though its manifestations vary widely—from full socioeconomic assimilation to persistent marginalisation (Alba and Nee, 2003; Portes and Rumbaut, 2001). Importantly, integration need not entail abandonment of one's origin culture. This recognition has motivated the development of multidimensional frameworks for understanding integration. In a seminal contribution, Berry (1980) analyses commitment to both origin and destination countries across multiple socioeconomic dimensions. Building on this, Constant and Zimmermann (2008) and Constant et al. (2009) introduce the Ethnosizer, a multidimensional measure of integration rooted in Berry's two-dimensional framework.

In this paper we make two distinct contributions. First, on the *empirical* side, we transform the Ethnosizer from a static, cross-sectional index into a fully *dynamic*, transition-based tool. Using longitudinal data from the German SOEP (2008–2019), we construct domain-specific Ethnosizers in national identity, media use, and social interaction. We estimate their evolution with dynamic discrete-choice methods —more specifically, a simultaneous dynamic binary-response model with correlated random effects. This allows us to quantify persistence and mobility in the identity states defined by Berry (1980), namely, integration, assimilation, separation, and marginalisation, and to document when and how individuals move between them.

Second, on the *theoretical* side, we develop a two-period model of identity choice with regret and rejoicing. In the model, immigrants can undertake a risky investment in integration based on their own expectations. Realized outcomes are compared to counterfactuals, generating ex post regret or confirmation; these retrospective evaluations can trigger revisions of identity, including what we term *integration reversal*, where immigrants step back from previously integrated positions toward more separated ones.

Relative to Constant et al. (2009), our contribution is therefore twofold. We move beyond the cross-sectional use of the Ethnosizer by modelling the *joint* dynamics of identity states across multiple domains and generating transition matrices. In addition, we explicitly link these empirical dynamics to an investment-and-regret framework

that provides an economic interpretation of both persistence and reversals in identity trajectories. To our knowledge, no prior work has treated the Ethnosizer as a dynamic state variable embedded in a structural transition framework and connected directly to a model of identity revision.

Our perspective is explicitly economic. Following Baumeister (1986) we model identity as a belief on the self. In our framework, this belief is revised after ex-post evaluation of the outcome of an integration investment. Misalignments between expectations and the realized returns to migration and integration can lead lead individuals to regret their previous choices and revise their identity. By combining panel data with up-to-date dynamic discrete-choice tools, our analysis quantifies the extent of such revisions and shows how they vary across domains.

Actually, using SOEP panel data, we find that commitments to destination and origin cultures are highly persistent once formed, yet not irreversible. Transition probabilities reveal substantial stability alongside non-negligible moves between identity states, including transitions towards less integrated states. These patterns align with the mechanisms of our identity-and-regret model: disappointing or favorable realizations of socioeconomic outcomes can lead to identity revision, consolidating or weakening the beliefs.

Taken together, our empirical and theoretical results highlight that identity formation is a dynamic, path-dependent process rather than a one-off outcome. The dynamic Ethnosizer and transition-based analysis provide a quantitative characterization of these trajectories, while the theoretical model offers an economic rationale for observed persistence and reversal.

The paper proceeds as follows. Section 2 reviews the essential literature on immigrant integration and identity and presents the construction of the Ethnosizer. Section 3 presents the SOEP data and the descriptive evidence. Section 4 is the core of our empirical analysis. It develops and estimates the correlated random-effects dynamic binary-response model that illustrate transitions across Berry states. Section 5 introduces our identity-and-regret model. Section 6 concludes.

2 Literature and theoretical background

According to a well-known formulation by Baumeister (1986), "an identity is a definition, an interpretation of the self". In game-theoretic terms, an interpretation of the self can be formalized as a belief. Because identhity implies a sense of belonging to a specific social group -characterized, for instance, by a common religion, language, and tradition (Chandra, 2012; Minoru Masuda and Meredith, 1970; Makabe, 1979; Unger et al., 2002; Phinney, 1990)- it follows that identity can be modeled as a belief in belonging to such a social group. Bénabou and Tirole (2011) go a step further by recognizing that this sense of belonging -this identity- generates utility and can thus be treated as an asset. Given that identity requires social recognition, its returns depend on one's reputation within their own group. This consideration has two consequences: 1) individuals invest in order to enter groups, enhance and maintain their status; 2) the returns to these investments are *uncertain*.

In plural societies, there exist different social groups with different identities. As a consequence, identity is to some extent a choice among different assets. This is especially true for immigrants, who are a minority and face the uncertainty of adaptation to a new social context.

According to Akerlof and Kranton (2000), "choice of identity may be the most important economic decision people make". However, important as it is, we point out that this decision *can be revised*. People can change their religion, their citizenship, or their way of life. In addition, exogenous shocks may also modify the self-image -that is, their identity. This is the case with a loss or gain of social status, or even health shocks.[1] *Changes* in identity and, in particular, its *revisability* have been little investigated in both the empirical and theoretical literature.[2] They are the focus of our contribution.

[1] Carillo et al. (2026) study a racially motivated attack carried out by an organized crime group in Italy against the local community of African immigrants. They find that the commitment of African immigrants to Italy increased in areas where the massacre prompted demonstrations of solidarity by Italians.

[2] In other disciplines, like psychology, the dynamic approach to identity analysis is more developed. Runge and Maehler (2026), for instance, study the trajectories of the cultural identity of immigrants in Germany according to a Berry-type model. Their findings are in line with our results, indicating that identity is persistent, but variable.

We empirically document the changes and revisability in identity in Sections 3-4. Then, in Section 5, we develop a theoretical model offering a possible mechanism of identity revision.

The literature provides two main frameworks to explore the concept of identity among immigrants; namely, the one-dimensional and the two-dimensional model. The first suggests that adaptation to the host society requires dismissing the original identity Herberg (1955), implying that there exists a negative correlation between a measure of commitment to the Origin country and a measure of commitment to the Destination country.

This assumption is somewhat arbitrary, and does not capture other possibilities that immigrants have in forming and modifying their identity. The two-dimensional model acknowledges that individuals can adapt to the host society without dismissing their original identity. Formally, this implies that the measures of commitment to the Origin and Destination country can be stochastically dependent, but not deterministically like in the one-dimensional model, where the degree of commitment to one country perfectly informs about the commitment to the other country[3]. In this approach, pioneered by Berry (1980), immigrants can be classified into four states:

- Assimilated: immigrants fully adopt the customs, values, and behaviors of the natives. They dismiss, in part or in whole, their original identity.
- Integrated: Immigrants adhere to the host country identity while preserving their original one.
- Separated: Immigrants preserve their original identity adopting little or none from the host society.
- Marginalised: Immigrants reject both the origin and the host identity.

[3]In other words, commitment to the origin and destination countries is determined independently, as we do in Section 5. This rules out direct spillovers from one dimension to the other, but it does not imply that the two dimensions are uncorrelated in the data. Correlation, whether positive or negative, may occur; the theory, however, does not impose a trade-off between commitment to the origin country and commitment to the destination country.

For a more comprehensive understanding of the two models, it is worthwhile to examine the approach outlined by Constant et al. (2009). Let us consider the level of commitment to both the host country and the origin country as two continuous variables: $Y, X \in [0, 1]$. Within this range, zero denotes the minimum level of commitment, and one the maximum level. In the one-dimensional model, the relationship between these variables is a simple linear function: $Y = 1 - X$. Figure (1) illustrates this approach, where an increase (decrease) in the commitment to the origin country results in a simultaneous decrease (increase) in the commitment to the host country.

The two-dimensional approach allows us to divide the plane into four quadrants. Figure (2) illustrates the "identity states" outlined by Berry (1980). In the top-left quadrant, the individual is *assimilated*, showing a strong commitment to the host country and a low commitment to the country of origin; in the top-right quadrant the individual is *integrated*, showing commitment to both countries; in the bottom-right quadrant the individual is *separated*, showing commitment only to the country of country; in the bottom-left quadrant the individual is *marginalised*, showing no commitment to either country.

Intertemporal movements across Berry's identity states are the core of our analysis. Our primary aim is to find empirical evidence that transitions between these states are empirically relevant, and to provide a theoretical foundation for them.

A major result of our empirical analysis concerns "backward moves" —transitions from more integrated states towards less integrated ones. To simplify the interpretation of our findings, we illustrate here our notion of "integration reversal".

Consider a social planner in the Destination country, and suppose that she prefers immigrants who are highly committed to Destination. She ranks Berry's identity states as follows:

$$\text{Assimilated} \succ \text{Integrated} \succ \text{Separated} \succ \text{Marginalized}.$$

We define integration reversal a situation in which an individual's identity in period $(t+1)$ moves from a more preferred state to a less preferred state (e. g., from Integrated to Separated). Notice that the ranking does not require a strict preference for assimilation over integration, nor for separation over marginalisation. The social planner may well be indifferent between integrated and assimilated immigrants, as well as equally

concerned about separated and marginalized ones. To detect integration reversal, we only need a strict preference between integrated and separated —that is, the planner must strictly prefer well-integrated immigrants over poorly-integrated ones. However, collapsing Berry's four states into only two categories would waste useful information. We therefore adopt a strict ordering to capture as many cross-states moves as possible.

To detect transitions across Berry states, we use the measure of identity introduced by Constant and Zimmermann (2008) and further developed by Constant et al. (2009); i. e., the Ethnosizer. This approach translates Berry's (1980) bidimensional identity framework into an operational measure at the individual level. We describe its construction in the following section.

2.1 The construction of the Ethnosizer

In order to build the Ethnosizer, Constant and Zimmermann (2008) and Constant et al. (2009) observe several proxies across five domains — language, national identity, culture, migration history, and social interaction — and classify each response according to the two relevant dimensions: commitment to the country of origin and commitment to the country of destination. Crossing these two dimensions yields the four identity states identified by Berry: assimilation, integration, separation, and marginalisation. For each proxy, they assign the individuals a score from 0 to 5, obtained by summing the positions across the five domains. For example, if an individual is separated in the linguistic domain, integrated in media use, assimilated in the national identity and social interaction domains, marginalized in migration history, she has a score of 1 in separation, 1 in integration, 2 in assimilation, and 1 in marginalisation. By construction, the total score adds up to five. As stated above, in our dynamic implementation we consider the *changes* along each dimension of the Ethnosizer. Specifically, we estimate a simultaneous dynamic binary-response model with correlated random effects, which allows destination- and origin-country commitment to evolve jointly while accounting for persistent individual heterogeneity. The model delivers transition probabilities across the four Berry states and therefore maps directly into the dynamic object that is central to our analysis.

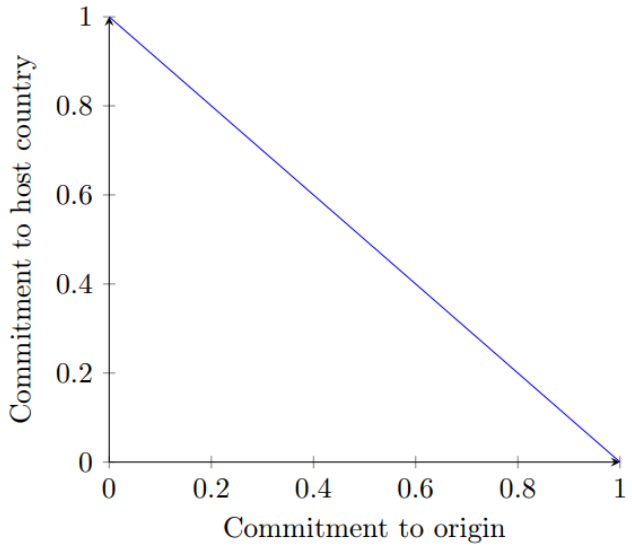


Figure 1: One-dimensional model

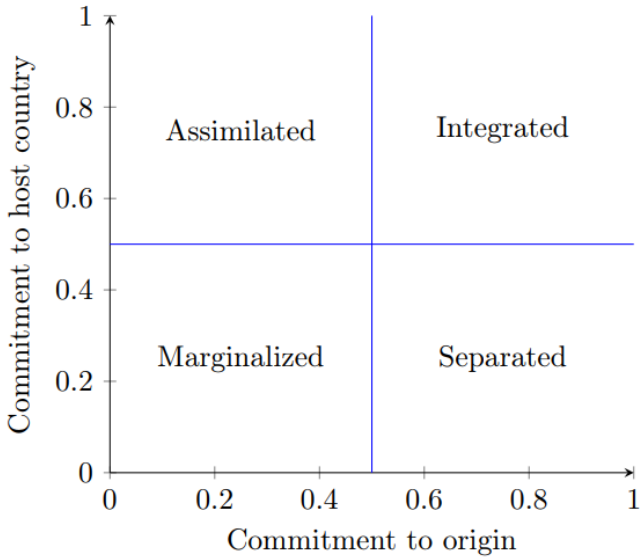


Figure 2: Two-dimensional model

3 Data and descriptive evidence

Consistently with our dynamic implementation, we exclude language from the Ethnosizer domains. Its inclusion is well motivated in the static Ethnosizer, where language helps characterize an immigrant's acculturation profile at a particular point in time. In a dynamic setting, however, language proficiency is less appropriate as a transition indicator because it is a slow-moving and cumulative form of human capital. For immigrants, language skills typically improve through prolonged exposure to the host-country environment, schooling, employment, and routine social interaction, rather than through discrete changes in identity orientation. Including language in the dynamic Ethnosizer would therefore risk attributing gradual and partly mechanical language acquisition to meaningful transitions across identity states. Hence, we compute the Ethnosizer with respect to three indicators; namely, self-identification with the country of origin/destination, media use, and social interaction. We work with SOEP-CORE waves 2008–2019; earlier years are dropped so all survey instruments are comparable.

National identity. Respondents rate how German they *feel* and how *committed* they are to their country of origin (1=*very strongly*, 5=*not at all*). Each item ≤ 3 is interpreted as a strong identification, yielding $g = 1$ for *Feeling German* and $o = 1$ for *Attached to Origin*.

Media use. Before 2012 the survey asked in which language respondents read news (`ReadsNewsOriginDE`); from 2012 onward it asked their preferred TV language (`Media Language`). We harmonise them into one variable by coding categories "German only" or "mixed" as $g = 1$ and categories "Origin only" or "mixed" as $o = 1$.

Social interaction. The survey records the frequency with which respondents visit or are visited by Germans and non-Germans during the previous year (1 = often, 4 = never).

We average the two items concerning contact with Germans; a mean below 2 indicates

frequent contact ($g = 1$).

The same rule is applied to the two items concerning contact with non-Germans, yielding $o = 1$.

Numeric outcomes 1–3. For each domain we derive two *binary* indicators, "German" ($g = 1$) and "Origin" ($o = 1$), and then cross them to obtain an unordered outcome with four categories, i.e. states ($s \in \{1, 2, 3, 4\}$), according to the following mapping:

$$(g, o) = \begin{cases} s = 1: & (1, 1) \text{ Integration}, \\ s = 2: & (1, 0) \text{ Assimilation}, \\ s = 3: & (0, 1) \text{ Separation}, \\ s = 4: & (0, 0) \text{ Marginalisation}. \end{cases}$$

The resulting numeric variables are named `ethnicity_acc`, `media_acc`, and `social_acc`, each taking values 1 to 4 and carrying the common label set

$$1 = \text{Integration},\ 2 = \text{Assimilation},\ 3 = \text{Separation},\ 4 = \text{Marginalisation}.$$

Table 1: Descriptive statistics of Ethnosizer scores

	Mean	SD	Min	Max
Assimilation score	.9279493	.7998367	0	3
Integration score	1.677487	.7651824	0	3
Separation score	.3106361	.5802499	0	3
Marginalisation score	.0839272	.2875961	0	3

Table 1 reports summary statistics for the Ethnosizer identity scores. Integration emerges as the dominant component, with an average score of about 1.68. This suggests that the most common profile is one in which commitment to the host and origin

contexts coexists rather than being mutually exclusive, supporting the bi-dimensional approach. Assimilation is the second most prevalent component, with a mean close to 0.93. This indicates that a non-negligible share of the observed domain-specific orientations is tilted mainly toward the host society. Separation is less common, with an average score of about 0.31, and therefore appears to be a comparatively less frequent configuration in the sample.

Marginalisation is the least prevalent component, with a mean of about 0.08. Although clearly less common than the other three components, it is not absent from the data: the standard deviation is around 0.29 and the maximum reaches 2, indicating that some individuals do display marginalisation in at least one, and occasionally two, of the domains. This suggests that disengagement from both origin and host contexts is relatively uncommon, but still empirically relevant.

Overall, the distribution of Ethnosizer score components is markedly unbalanced. Most of the mass is concentrated in integration, followed by assimilation, while separation is less frequent and marginalisation remains comparatively limited. Taken together, these descriptive patterns suggest that individuals in the sample are more likely to combine commitment to origin and host societies, or to lean toward the host society, than to occupy more detached or polarized configurations.

However, our aim is to investigate the possible moves among the Berry states over time, and prove that they are empirically relevant. In what follows, we document whether they change, how often they change, and whether the observed dynamics differ across domains.

The domain-specific perspective is important, because different domains of acculturation may evolve at different speeds and capture different aspects of immigrant incorporation. National identity reflects self-perceived belonging to the origin and destination country. Media use captures cultural and informational orientation. Social interaction captures interpersonal contact and network composition. Treating these domains separately allows us to assess whether the dynamics of acculturation are similar, or whether some domains are more persistent or more mobile than others.

3.1 Panel structure

A first important feature of the data is that the panel is substantially unbalanced. Table 2 reports the observation structure.

Table 2: Sample and panel structure

Statistic	Overall	National identity	Media use	Social Interaction
Individuals	1031	745	815	839
Obs. ≥ 2	703	618	616	669
Mean obs. ≥ 2	5.81	6.28	6.21	6.00
Median obs. ≥ 2	5	6	6	5
Mean year-to-year intervals		5.19	5.11	4.89
Median year-to-year intervals		5	5	4
% with year-to-year intervals		59.36	59.17	64.50

Overall, the panel includes 1,031 individuals, of whom 703 are observed at least twice. Coverage varies across domains. National identity is observed at least once for 745 individuals and at least twice for 618 individuals. Media use is observed at least once for 815 individuals and at least twice for 616 individuals. Social interaction has the broadest coverage, with 839 individuals observed at least once and 669 individuals observed at least twice. Among individuals with valid year-to-year intervals, the average number of intervals is very similar across domains, ranging from 4.89 for social interaction to 5.19 for National identity.

Figure 3 shows the distribution of the number of observed waves per individual. The figure highlights the unbalanced nature of the data: many individuals are observed in only one or two waves, while a smaller group is followed for a much longer period.

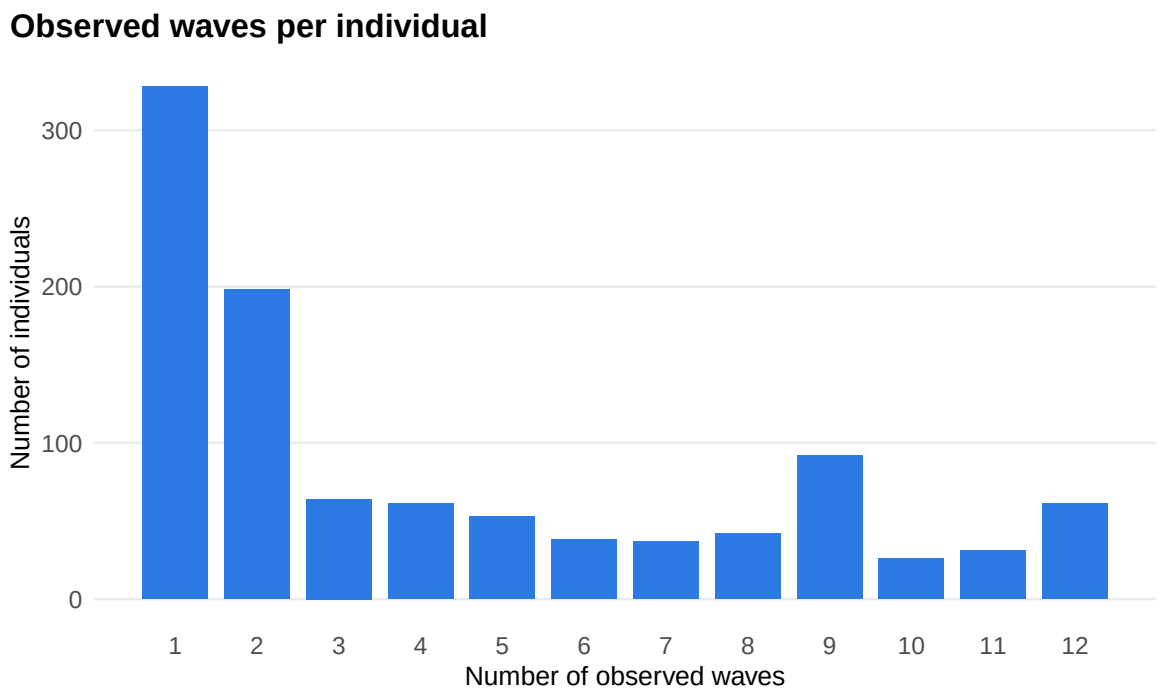


Figure 3: Observed waves per individual

This structure motivates the use of two complementary descriptive measures. First, we compute individual-level indicators, such as whether an individual ever experiences a transition and whether the first and last observed states differ. Second, we compute annual transition rates over valid year-to-year intervals, defined by consecutive non-missing observations for the same individual. The first approach captures whether mobility accumulates over individual histories, while the second describes short-run transition probabilities.

3.2 Distribution of Berry states

Table 3.2 reports the distribution of Berry states by domain. For each domain, we report the pooled distribution over all person-year observations, the distribution at the first observed non-missing state, and the distribution at the last observed non-missing state.

Table 3: Distribution of Berry states by domain

Domain	State	Pooled (%)	First obs. (%)	Last obs. (%)
National identity	Integration	59.09	58.52	62.28
National identity	Assimilation	24.81	22.01	22.15
National identity	Separation	15.06	17.72	14.09
National identity	Marginalisation	1.05	1.74	1.48
Media use	Integration	22.87	20.25	22.70
Media use	Assimilation	60.55	62.33	60.37
Media use	Separation	11.01	9.45	10.80
Media use	Marginalisation	5.57	7.98	6.13
Social interaction	Integration	84.44	83.55	82.24
Social interaction	Assimilation	8.22	8.70	8.10
Social interaction	Separation	5.19	5.48	6.20
Social interaction	Marginalisation	2.15	2.26	3.46

Several features stand out from Table 3.2. First, the distribution of Berry states differs sharply across domains. National identity is predominantly integrated: 59.09 percent of pooled observations fall into integration, while assimilation accounts for 24.81 percent and separation for 15.06 percent. Marginalisation is almost absent, at just 1.05 percent. This suggests that, when identity is measured as self-perceived belonging, most immigrants combine commitment to Germany with commitment to the origin country rather than replacing one with the other.

Second, media use displays a substantially different profile. Here, assimilation is the dominant state, accounting for 60.55 percent of pooled observations, while integration accounts for only 22.87 percent. This indicates that media consumption is much more destination-oriented than National identity. In other words, immigrants may retain a dual sense of belonging while consuming media primarily in the destination-country language. This difference is important because it suggests that behavioral and subjective dimensions of identity need not move together.

Third, social interaction is by far the most integration-oriented domain. In the pooled

distribution, 84.44 percent of observations are classified as integrated, while assimilation, separation, and marginalisation remain comparatively small. This pattern suggests that, at least in terms of social visits and contacts, many immigrants maintain ties with both Germans and foreigners/origin-related networks. Social interaction therefore appears less polarized than media use and more strongly concentrated in the integrated state.

The comparison between first and last observed states provides a first indication of medium-run changes in the composition of states. In National identity, integration increases from 58.52 to 62.28 percent, while separation declines from 17.72 to 14.09 percent. In media use, integration also increases slightly, from 20.25 to 22.70 percent. Interestingly, within this domain, assimilation declines from 62.33 to 60.37 percent. By contrast, social interaction is highly stable in the aggregate: integration remains above 80 percent at both the first and last observation, although it declines slightly from 83.55 to 82.24 percent and marginalisation rises from 2.26 to 3.46 percent.

Taken together, these patterns indicate that the three domains capture distinct dimensions of the identity. National identity tends to be predominantly dual, media use is largely oriented toward the destination culture, and social interaction is overwhelmingly integrated. This heterogeneity is substantively important: it suggests that there is no single empirical trajectory of identity formation, and that aggregating them into a single measure could result in a loss of valuable information.

Table 4: Cross-domain association between Berry states: Cramer's V

Domain	National identity	Media use	Social interaction
Identity	1.000	0.190	0.086
Media	0.190	1.000	0.131
Social	0.086	0.131	1.000

Table 4 reports Cramer's V for the association between Berry states across domains, using person-year observations. The values are positive but relatively small: 0.190 for ethnicity and media use, 0.086 for ethnicity and social interaction, and 0.131 for media

use and social interaction. This indicates that the three domain-specific classifications are related, but only weakly so. In substantive terms, knowing an individual's Berry state in one domain provides limited information about their state in another domain in the same year. This reinforces the domain-specific interpretation of the descriptive evidence.

Table 5: Summary dynamic indicators by domain

Domain	National identity	Media use	Social interaction
Year-to-year intervals	3177	3118	3255
Annual transition rate (%)	8.69	8.53	6.18
Individuals	612	610	665
Ever transition (%)	29.90	29.51	18.35
First-last change (%)	21.68	25.16	13.15

Considering whether individuals move across Berry states over time, table 5 reports aggregate dynamic indicators for the three domains. A valid year-to-year interval is defined by two consecutive non-missing observations for the same individual. A Berry-state transition occurs when the state differs between the beginning and the end of the interval. The annual transition rate is therefore the share of valid intervals containing such a state change. The table also reports the number of individuals contributing at least one valid interval, the share who ever experience a transition and the share whose first and last observed states differ.

The full raw annual transition matrices and their descriptive interpretation are reported in Appendix A. These matrices provide a benchmark for the model-based transition probabilities estimated in the following subsection.

4 Estimation of transition matrices

4.1 Simultaneous dynamic binary-response model with correlated random effects

We model the joint dynamics of two binary outcomes observed for each individual i across survey waves t.

The first indicator, $y_{1,i,t} \in \{0,1\}$, captures commitment with the destination-country dimension.

The second indicator, $y_{2,i,t} \in \{0,1\}$, captures commitment with the origin-country dimension in the national identity and media-use domains, and social interaction with non-Germans in the social-interaction domain.

The two indicators jointly define the four domain-specific states: Integration $(1,1)$, Assimilation $(1,0)$, Separation $(0,1)$, and Marginalisation $(0,0)$.

At each t an individual occupies one of four states $s_{i,t} \in \{1,2,3,4\}$, mapped as

$$s_{i,t} = 4 - 2y_{1,i,t} - y_{2,i,t} \quad \Longleftrightarrow \quad \begin{cases} s = 1: & (1,1) \text{ Integration}, \\ s = 2: & (1,0) \text{ Assimilation}, \\ s = 3: & (0,1) \text{ Separation}, \\ s = 4: & (0,0) \text{ Marginalisation}. \end{cases} \tag{4.1}$$

Our objective is to recover a 4×4 matrix of model-implied transition probabilities.

Rows correspond to states at $t-1$ and columns to states at t.

The matrices are obtained by averaging the model-implied probabilities over the individuals contributing observed transitions from each origin state, weighting each individual by the number of such transitions.

The exact construction of the estimated matrices is provided in Appendix B, and, more specifically, in Equation (B.8).

4.2 Results

This section presents the estimated transition matrices for first-generation immigrants across three domains: media use, social interaction, and national identity. Each 4×4 matrix reports the predicted probability of transitioning from the origin state at $t-1$ (row) to the destination state at t (column). Rows sum to approximately one by construction. Sankey diagrams are used to visualize the estimated transitions in the main text, while the numerical values are reported in Table 7 at the end of this section.

Table 6: Summary of transition dynamics by domain

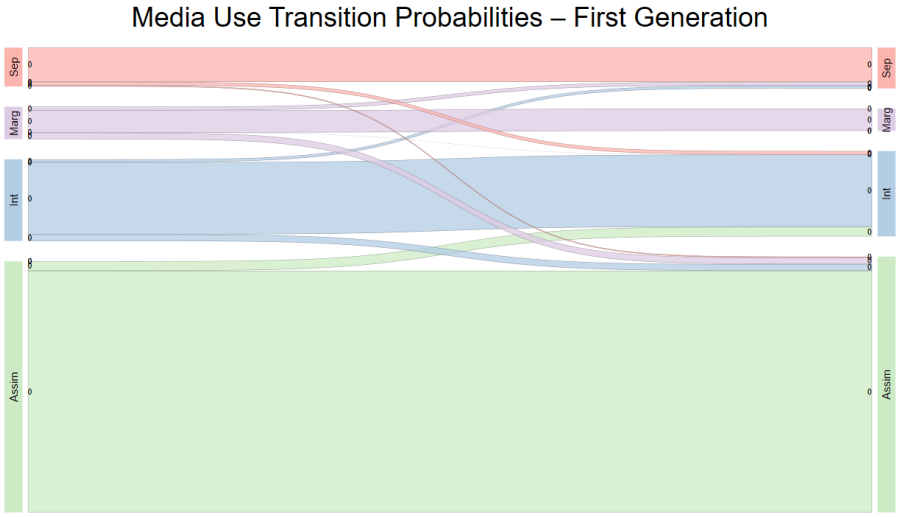
Domain	Mean persistence	P(int. from marg.)	P(int. from sep.)	P(exit from int.)
Media use	0.842	0.003	0.089	0.116
Social interaction	0.804	0.196	0.177	0.035
National identity	0.812	0.016	0.151	0.063

Note: Mean persistence is the unweighted average of the four diagonal entries in each transition matrix. Probabilities into integration are taken from the marginalisation and separation rows; probabilities of exiting integration are defined as $1 - P(\text{integration} \rightarrow \text{integration})$.

Table 6 distils the main patterns from the full transition matrices. Across all three domains, persistence is high: the average diagonal entry lies around 0.80–0.84 confirming that identity states are strongly state-dependent. Moves into integration from marginalisation are rare in media use and national identity (below 0.02), but much more frequent in social interaction. Transitions from separation into integration are present in all domains, with estimated probabilities of 0.089 in media use, 0.177 in social interaction, and 0.151 in national identity. At the same time, the probability of exiting an integrated state is small but non-negligible—between 3% and 12% depending on the domain. These figures underscore that integration is both sticky and revisable: most individuals remain in their current state from one period to the next, yet a non-trivial share either climbs into integration or slips out of it, with the balance of these flows varying systematically across domains.

Media use

Consistent with Table 6, media use displays high persistence, with mean diagonal probabilities around 0.83–0.84. Once individuals settle into a given media-use state, especially assimilation or integration, they tend to remain there. Moves from marginalisation directly into integration are rare, whereas transitions from separation into integration are more common (about 9%). Table 8 shows that reverse movements are nevertheless present. The most important are transitions from Assimilation to Integration (3.7%) and from Integration to Separation (3.5%), while transitions toward Marginalisation are less frequent.



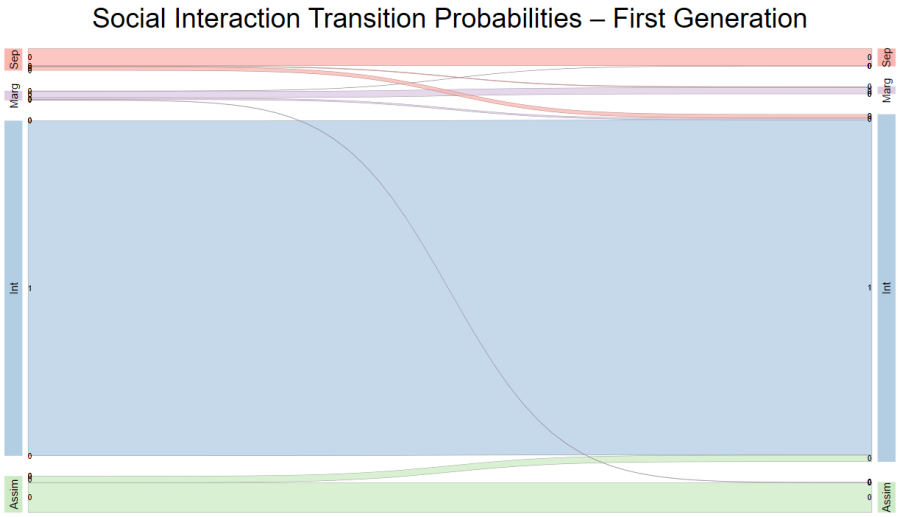
(a)

Figure 4: Estimated transition probabilities: media use

Social interaction

Social interaction also features strong state dependence, with mean persistence close to 0.80 (Table 6). Upward moves into integration are frequent: nearly one in five marginalised individuals and a similar share of separated individuals transition into integration from one period to the next. The probability of exiting an integrated state is very low (approximately 3.5%), indicating that, once regular social interaction with

both Germans and non-Germans is established, it is rarely abandoned. Table 8 shows that reversals from Integration are limited: the probabilities of moving from Integration to Separation and Marginalisation are 1.3% and 0.5%, respectively. The largest reverse movement in this domain is instead from Assimilation to Integration, at 18.3%.

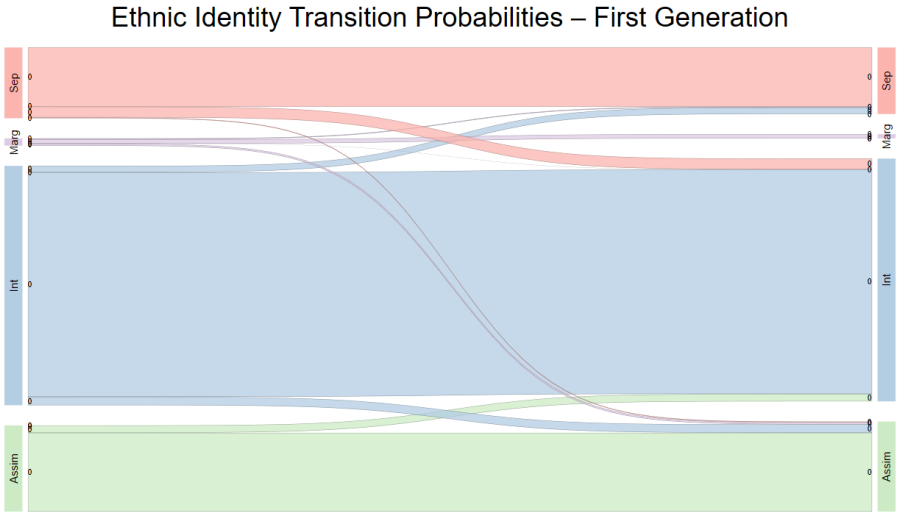


(a)

Figure 5: Estimated transition probabilities: social interaction

National identity

For national identity, persistence is again high (mean diagonal 0.81). Transitions from marginalisation into integration are rare, but moves from separation into integration are more common (approximately 1.6% from marginalisation and 15.1% from separation), indicating that "one-sided" national profiles can evolve into more balanced ones over time. Reverse movements are also observed. As shown in Table 8, the largest is from Assimilation to Integration (8.5%), while Integration to Separation occurs with probability 2.7%. Transitions from Assimilation to Separation or Marginalisation and from Separation to Marginalisation are below 1%. Together, the national-identity dynamics show that identity is sticky, but revisable.



(a)

Figure 6: Estimated transition probabilities: national identity

Summary Across all three domains, the estimated dynamics point to strong state persistence with gradual mobility. Using mean diagonal persistence as the criterion, media use is the most stable domain (0.842), followed by national identity (0.812) and social interaction (0.804). This ranking should not, however, be interpreted as an equivalent ranking of integration reversal. As Table 8 shows, the incidence of reverse movements depends strongly on the origin state and on the particular transition considered.

Table 7: Estimated transition probabilities by domain (row percentages)

Domain	State at t-1	Integr.	Assimi.	Separ.	Margin.
National identity	Integration	93.70	3.50	2.70	0.00
National identity	Assimilation	8.50	90.40	0.40	0.70
National identity	Separation	15.10	1.10	82.80	0.90
National identity	Marginalisation	1.60	28.30	12.10	58.00
Media use	Integration	88.40	8.10	3.50	0.00
Media use	Assimilation	3.70	93.90	0.70	1.70
Media use	Separation	8.90	2.70	86.10	2.30
Media use	Marginalisation	0.30	19.90	11.40	68.40
Social interaction	Integration	96.50	1.60	1.30	0.50
Social interaction	Assimilation	18.30	79.40	0.00	2.30
Social interaction	Separation	17.70	0.00	78.20	4.00
Social interaction	Marginalisation	19.60	7.60	5.30	67.50

Table 8: Integration reversals by domain (row percentages)

Transition	National identity	Media use	Social interaction
Assimilation $\rightarrow$ Integration	8.50	3.70	18.30
Assimilation $\rightarrow$ Separation	0.40	0.70	0.00
Assimilation $\rightarrow$ Marginalisation	0.70	1.70	2.30
Integration $\rightarrow$ Separation	2.70	3.50	1.30
Integration $\rightarrow$ Marginalisation	0.00	0.00	0.50
Separation $\rightarrow$ Marginalisation	0.90	2.30	4.00

Notes: The table reports model-implied probabilities of transitions toward lower-ranked Berry states, according to the ordering Assimilation $\succ$ Integration $\succ$ Separation $\succ$ Marginalisation introduced in Section 2. Each entry is the probability of the indicated destination state conditional on the origin state.

5 A two-period model of identity formation

5.1 The model

In the previous sections, we have shown that identity, though persistent, can change over time and, moreover, that it may revert from more integrated states to less integrated ones. Here, we provide a novel theoretical model of identity formation that accounts for these phenomena. We consider a two-period model in which, for simplicity, there are only two levels of commitment to a country: low and high. The immigrant is uncertain about her own levels of commitment (namely, her type): upon arriving in the *D*estination country (henceforth D), she does not really know how "nostalgic" she will feel toward her *O*rigin country (henceforth O), nor how well she will fit into the new social environment. She only has some beliefs about her commitment to O and D. These beliefs constitute her self-image—namely, her identity.

In the first period, the immigrant can make a risky investment in integration. In the second period, the returns to the investment are realized and the beliefs about the commitment to O and D are revised, depending on the regret/rejoicing in the choices made in period 1.

Formally, let α (resp. β) denote the *low* level of commitment to O (resp. D). Analogously, let A (resp. B) denote the *high* level of commitment to O (resp. D). For each agent i and each period $t = 1, 2$, we let $p_{i,t}$ denote the probability assigned to type B, the high level of commitment to D, and $(1 - p_{i,t})$ the probability assigned to type β. Similarly, we let $q_{i,t}$ denote the probability assigned to type A, the high level of commitment to O, and $(1 - q_{i,t})$ the probability assigned to type α.

The individual identity in period t is then a joint probability distribution on $\{\alpha, A\} \times \{\beta, B\}$, namely:

belief on (α, β) :	$(1 - p_{i,t})(1 - q_{i,t})$	(marginalized)
belief on (A, β) :	$q_{i,t}(1 - p_{i,t})$	(separated)
belief on (α, B) :	$p_{i,t}(1 - q_{i,t})$	(assimilated)
belief on (A, B) :	$p_{i,t}q_{i,t}$	(integrated)

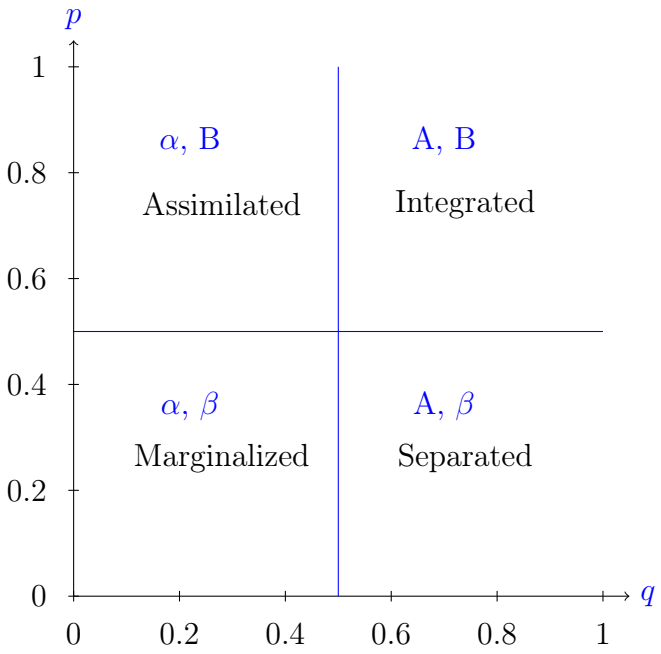


Figure 7: Identity states in the (q, p) plane.

This list provides a simple tool to represent identity states. We proceed as follows. First, we construct figure 7 placing p_{it} on the vertical axis and q_{it} on the horizontal axis. Then, we define the cutoff values $p_{i,t}, q_{i,t} = 1/2$, and assume that an individual i is of type A (resp. B) when $q_{i,t}$ (resp. $p_{i,t}$) $> 1/2$.[4] This divides the figure into four squares associated with the two levels of commitment to O and D; namely, $(\alpha, \beta); (A, \beta); (\alpha, B); (A, B)$. These squares reproduce the Berry (1980) model as follows: (α, β) stands for Marginalized, (A, β) for Separated, (α, B) for Assimilated, and (A, B) for Integrated. In the figure, each pair $(q_{i,t}, p_{i,t})$ falls into one square.[5]
We associate each pair $(q_{i,t}, p_{i,t})$ with the Berry state determined by whether each marginal belief lies above or below $1/2$. Equivalently, the prevalent identity is the state receiving the largest probability mass, except at knife-edge values at which two or more states receive the same mass.

Through this rule, we assign a prevalent identity to any pair $(q_{i,t}, p_{i,t})$, and provide a

[4] Conversely, she is of type α if $q_{i,t} \leq 1/2$ and of type β if $p_{i,t} \leq 1/2$.
[5] Figure 7 can be interpreted as an Edgeworth box.

theoretical foundation for the Berry (1980) model.

Now we turn to the investment in integration. Upon arrival in D, immigrants can invest time, effort, or money to improve their social integration. This investment is a risky one, as its returns are uncertain —the integration effort may prove unsuccessful. Additionally, we note that in social adaptation processes, the uncertainty concerns not only the returns, but also the costs. Indeed, when adopting new behaviors, no one truly knows whether they will be able to comply with them. In other words, costs vary significantly according to individual characteristics and are revealed only over time. This fact adds further uncertainty to investing in integration. As a result, immigrants with "pessimistic" beliefs may find the investment disadvantageous. On the other hand, not investing is a safe action that has no costs.[6] For simplicity, we assume that the *net* return to the investment takes only two values ("low" or "high"), and that the return to the safe action lies between "low" and "high".

In the second period, the net returns are observed and individuals assess their first-period choice. Formally, they compare the *realized* utility under the action actually taken with the counterfactual utility they would have obtained under the alternative action. The counterfactual is observable, for example, by looking at peers. If the realized utility falls short of the counterfactual benchmark, they experience ex post regret; if it exceeds it, they experience ex post rejoicing. We assume that regret and rejoicing may induce a revision of the initial beliefs on the commitment level to O and Destination —namely, on the identity.

Note that the ex post comparison introduces a *retrospective* dimension into identity formation. In other words, agents re-evaluate their past choices in the light of their consequences. Regret and rejoicing arise as standard ex post utility comparisons. We capture this behavior by allowing second-period identity to differ from the first period one. Below, we develop a theoretical framework that intuitively describes this situation.

[6] We can easily introduce uncertainty in the returns from not investing. However, this would only complicate the cutoff rule that separates investors from non-investors, and we prefer to preserve simplicity.

5.2 The investment in integration

In period $t = 1$, each individual $i = 1, 2, ..., n$ chooses an action $a_i^1 \in \{0, 1\}$, where 0 means "not investing" and 1 means "investing".
The set of *net* returns on the investment is $\Theta = \{\theta_H, \theta_L\}$. The return θ_S to the safe action $a_i^1 = 0$ is certain. We assume

- A_1: for all i, $u_i(0, \theta_L) > u_i(1, \theta_L)$.
- A_2: for all i, $u_i(0, \theta_H) < u_i(1, \theta_H)$.

Assumptions A1 and A2 imply $\theta_L < \theta_S < \theta_H$. In other words, investors who receive the low payoff (low net return) are worse off than non-investors[7] Investing is optimal when the following decision rule is satisfied:

$$a_i^1 = 1 \iff \mathbb{E}_{\mu_i}\left[u_i(1, .)\right] \geq \mathbb{E}_{\mu_i}\left[u_i(0, .)\right]. \tag{5.1}$$

where $u_i(., .)$ is the VNM individual utility function, and $\mathbb{E}_{\mu_i}\left[u_i(., .)\right]$ the expectation according to the probability measure μ_i. Equation 5.1 can be rewritten as

$$u_i(1, \theta_L)\mu_i(\theta_L) + u_i(1, \theta_H)\mu_i(\theta_H) \geq u_i(0, \theta_L)\mu_i(\theta_L) + u_i(0, \theta_H)\mu_i(\theta_H), \tag{5.2}$$

or, equivalently—via simple algebraic steps—as

$$\mu_i(\theta_H) \geq m\, \mu_i(\theta_L), \tag{5.3}$$

where

$$m = -\frac{u_i(0, \theta_L) - u_i(1, \theta_L)}{u_i(0, \theta_H) - u_i(1, \theta_H)} > 0$$

Condition 5.2 means that, in order to invest, the beliefs about the high return must be sufficiently high. Consequently, all immigrants with "pessimistic" beliefs on the net return of the investment in integration choose the safe action.[8]

[7] For simplicity, in this formulation we disregard anticipated regret because we only want to show how immigrants self-select between investors and non-investors. Including anticipated regret would change the cutoff value of $\mu_i(\theta_L)$ without gains in generality.

[8] Condition 5.2 assumes for simplicity that when the expected return from investing and non-investing is the same, the individual invests. Notice also that $m > 0$. The numerator is positive by A_1, and the denominator is negative by A_2

5.3 Second Period: the Existential Life-Assessment

In the second period, $\theta \in \Theta$ is realized and observed. Thus, anyone can assess the consequences of their choice. Our approach to the existential life-assessment is grounded in the regret theory of Loomes and Sugden (1982, 1986). Regret/rejoice is defined as the difference between the *realized* utility of the first-period choice and the utility had the alternative first-period choice been made. Notice that in our framework there exist *two* sources of regret: one concerns the choice of investing or not in integration, the other concerns the choice to migrate. Migration decisions are particularly vulnerable to regret, and retrospective assessments are common among immigrants. For instance, survey evidence on a sample of immigrants in Italy we are collecting reveals that 30% of respondents declare to regret the decision to migrate.[9]

5.4 Regretting or Rejoicing the Integration Investment

Consider an immigrant who has chosen $a_i^1 = 0$. In our framework, she experiences regret if $\theta = \theta_H$, and rejoicing if $\theta = \theta_L$.
By contrast, an immigrant who has chosen $a_i^1 = 1$ experiences regret if $\theta = \theta_L$, and rejoicing if $\theta = \theta_H$.

5.5 Regretting or Rejoicing the Migration Decision

As mentioned above, immigrants may face another form of regret, namely, the one concerning their decision to migrate. In assessing their lives, they inevitably compare their realized outcomes to what they could have obtained in their origin country. In practice, this can be done by observing non-immigrant, comparable peers.
Residents in O receive a random revenue λ, which can be "low" or "high" —namely, $\lambda \in \Lambda = \{\lambda_H, \lambda_L\}$.

[9] When asked whether they regret the decision to migrate or the choice of Italy as their destination, 98% of these immigrants choose the first option (source: Next2EU database, GRINS Project).

We order the outcomes in D and O as follows:

$$\lambda_L < \theta_L < \lambda_H < \theta_H. \tag{5.4}$$

This ordering assumes that the low revenue in O represents the worst-case scenario, while the successful investment in D represents the "first best." To understand the economic intuition behind this assumption, suppose that O is "poor" and D is "rich." The ordering mirrors the stylized fact that economic downturns in poor countries tend to be particularly severe, while the returns on investments in rich countries are generally higher.[10] However, the point we wish to emphasize through the ordering (5.4) is that the high revenue in O (λ_H) may still be preferable to a failed investment (θ_L) in D. Additionally, we assume that the high revenue in O is preferable to the return from the safe action in D, namely $\theta_S < \lambda_H$.
Together, these assumptions capture the possibility that emigration is ex-post worthless, generating feelings of frustration and regret.
The above can be summarized as follows:

$$\lambda_L < \theta_L < \theta_S < \lambda_H < \theta_H. \tag{5.5}$$

In this framework, computing the regret concerning the emigration decision boils down to comparing the $\theta \in \Theta$ realized in D with the $\lambda \in \Lambda$ realized in O. Four cases are possible:

(i) (λ_L, θ_L)
(ii) (λ_L, θ_H)
(iii) (λ_H, θ_L)
(iv) (λ_H, θ_H).

We can study the possible effect of regret on identity in each case. Specifically, we argue that regret is a mechanism capable of generating a revision in identity. To clarify our argument, let us recall that, according to Baumeister (1986), the identity can be

[10] Actually, we do not need to assume $\lambda_H < \theta_H$, but we do so for sake of simplicity and without loss of generality.

understood as a belief about the self. In the immigrant identity, feelings of commitment to O and D play a pivotal role. In our approach, these feelings are summarized by $p_{i,t}$ and $q_{i,t}$; namely, the belief of being highly committed to D or O respectively. Consequently, a revision of these beliefs implies a revision of identity.
Below, we explain the direction in which regret tends to modify $p_{i,2}$ and $q_{i,2}$. We are aware that this is merely an initial attempt to formalize this mechanism, and given the lack of an established consensus, we propose three criteria that we find "natural":

1) if there is no regret of emigrating (i. e., the emigration choice proves to be correct), then $p_{i,2} \geq p_{i,1}$. If there is regret of emigrating (i. e., the emigration choice proves to be wrong), then $p_{i,2} \leq p_{i,1}$.

2) if there is no regret of investing (i. e., the investing choice proves to be correct), then $p_{i,2} \geq p_{i,1}$. If there is regret of investing (i. e., the investing choice proves to be wrong), then $p_{i,2} \leq p_{i,1}$.

3) If λ_L is realized, then $q_{i,2} \leq q_{i,1}$. If λ_H is realized, then $q_{i,2} \geq q_{i,1}$. This is done because a low revenue in O does not incentivize longing and homesickness; the opposite occurs when the high revenue is realized.

Note that, in considering the effect of the regret, we must take into account whether the immigrant has invested or not in the first period. Thus, each case i), ii), iii), iv) includes the sub-cases $a_i^1 = 0, 1$.

5.5.1 Case i): λ_L in the Origin and θ_L in the Destination.

$\boldsymbol{a_i^1 = 0)}$ By Criterion 1) and Criterion 2), $p_{i,2} \geq p_{i,1}$. By Criterion 3, $q_{i,2} \leq q_{i,1}$.
$\boldsymbol{a_i^1 = 1)}$ By Criterion 1), $p_{i,2} \geq p_{i,1}$. By Criterion 2, $p_{i,2} \leq p_{i,1}$. The net effect is ambiguous. By Criterion 3), $q_{i,2} \leq q_{i,1}$.

5.5.2 Case ii): λ_L in the Origin, θ_H in the Destination

$\boldsymbol{a_i^1 = 0}$**)** By Criterion 1), $p_{i,2} \geq p_{i,1}$. By Criterion 2), $p_{i,2} \leq p_{i,1}$. The net effect is ambiguous. By Criterion 3), $q_{i,2} \leq q_{i,1}$.
$\boldsymbol{a_i^1 = 1}$**)** By Criterion 1) and Criterion 2), $p_{i,2} \geq p_{i,1}$. By Criterion 3, $q_{i,2} \leq q_{i,1}$.

5.5.3 Case iii): λ_H in the Origin, θ_L in the Destination

$\boldsymbol{a_i^1 = 0}$**)** By Criterion 1), $p_{i,2} \leq p_{i,1}$. By Criterion 2), $p_{i,2} \geq p_{i,1}$. The net effect is ambiguous. By Criterion 3), $q_{i,2} \geq q_{i,1}$.
$\boldsymbol{a_i^1 = 1}$**)** By Criterion 1), $p_{i,2} \leq p_{i,1}$. By Criterion 2), $p_{i,2} \leq p_{i,1}$. By Criterion 3), $q_{i,2} \geq q_{i,1}$. Notably, this is the worst possible scenario for an immigrant; namely, regretting all the choices she made.

5.5.4 Case iv): λ_H in the Origin, θ_H in the Destination

$\boldsymbol{a_i^1 = 0}$**)** By Criterion 1), $p_{i,2} \leq p_{i,1}$. By Criterion 2), $p_{i,2} \leq p_{i,1}$. The net effect is ambiguous. By Criterion 3), $q_{i,2} \geq q_{i,1}$.

$\boldsymbol{a_i^1 = 1}$**)** By Criterion 1), $p_{i,2} \geq p_{i,1}$. By Criterion 2), $p_{i2} \geq p_{i1}$. By Criterion 3), $q_{i2} \geq q_{i1}$. This is one of the best possible scenarios for an immigrant: all choices proved correct, and there is no regret at all.

5.6 Reconsideration and integration reversal

The framework above provides a mechanism through which retrospective evaluation of past choices can generate changes in identity. More precisely, changes in $p_{i,2}$ and $q_{i,2}$ relative to their first-period values represent revisions in beliefs about commitment to the Destination and Origin countries, respectively.

In the spirit of Berry (1980), such revisions generate an observed change of identity

whenever they are sufficiently large to move the individual across one of the boundaries separating the four Berry states. Individuals whose beliefs lie close to one of these boundaries are therefore more likely to change observed identity following a given revision in beliefs.

6 Conclusions

We examined the dynamics of identity among immigrants in Germany across three main domains (national identity, media use, and social interaction) using a simultaneous dynamic binary-response model with correlated random effects. Three major results emerge. 1) Identity, understood as a position in one of Berry's states, is highly persistent. 2) However, it is not static and can change over time. 3) Such changes can move forward and backward, in the sense that, over time, immigrants can step back from previously integrated positions toward more separated ones, generating the phenomenon we label "integration reversal."

This result brings to light the reversibility of the integration process. We offer a theoretical interpretation of this reversal through a regret-based model of identity formation. Although anecdotal evidence of this phenomenon exists, our analysis represents—to the best of our knowledge—the first attempt to identify it both empirically and theoretically.

The reversibility of the integration process raises serious concerns in all societies where immigration is substantial. Immigrants typically exhibit the highest fertility rates in these societies. Their integration is key to the sustainability of welfare systems and retirement benefits. The creation of permanently disadvantaged, separated minorities would not only generate problems of public order, but would raise even more questions about the sustainability of the welfare state. Preventing phenomena of integration reversal is therefore a matter of the utmost importance. In this respect, it is natural to wonder about the "external validity" of our results. Germany is a country that provides high-quality public education, a generous welfare system, and low unemployment rates. All of these factors support integration. We can conjecture that less prosperous countries could be more exposed to integration reversal, but research is needed to con-

firm this intuition.

In addition, although we have used the Ethnosizer and the Berry classification, other measures of identity and other classifications are conceivable. We plan to develop our research in this direction.

Finally, we have emphasized the role of regret because, according to the survey data we are also collecting, we believe that retrospection is important for affirming identity. However, other mechanisms could also be relevant. We leave these avenues for future research.

A Raw annual transition matrices

Table 9: Annual transition matrices by domain (row percentages)

Domain	State at t-1	Integr.	Assim.	Separ.	Margin.
National identity	Integration	93.76	3.28	2.85	0.11
National identity	Assimilation	8.43	90.83	0.50	0.25
National identity	Separation	14.79	1.67	83.54	0.00
National identity	Marginalisation	9.68	9.68	3.23	77.42
Media use	Integration	88.33	7.88	3.38	0.42
Media use	Assimilation	3.34	94.49	1.32	0.85
Media use	Separation	8.62	3.74	87.36	0.29
Media use	Marginalisation	6.47	11.18	2.94	79.41
Social interaction	Integration	96.38	1.66	1.37	0.58
Social interaction	Assimilation	18.22	80.30	0.00	1.49
Social interaction	Separation	18.52	0.00	79.01	2.47
Social interaction	Marginalisation	13.33	6.67	3.33	76.67

Table 9 reports the full annual transition matrices. Each row gives the distribution of the state observed at time t, conditional on the state occupied in the previous wave. The diagonal entries therefore measure persistence, while the off-diagonal entries show the direction of the movements across Berry states. Berry states are highly persistent from one year to the next. Diagonal transitions are strongly over-represented relative to independence, while most off-diagonal movements are under-represented.

The table reinforces the main substantive point: persistence is high across domains, but it is anchored in different states. Social interaction is most clearly organized around integration, with 96.38 percent of integrated observations remaining integrated in the following wave. Media use is instead most strongly organized around assimilation, which has the highest persistence in that domain. National identity lies between these two cases, with high persistence in both integration and assimilation but a stronger

orientation toward integration than the media domain.

The off-diagonal entries also show that mobility is usually limited and state-specific rather than a wholesale reshuffling of positions. When movements occur, they tend to involve transitions toward the dominant state of the domain: integration in national identity and social interaction, and assimilation in media use. These patterns suggest that acculturation is not a single unified process unfolding at the same pace across domains, but a set of domain-specific trajectories with distinct dynamic equilibria.

It is useful to compare these raw transition frequencies with the model-based probabilities reported in Section 4. The two sets of results are remarkably similar for the most frequently observed states. In particular, the persistence of Integration is virtually unchanged across the two approaches, and transitions from Separation into Integration are also very close in all three domains. The same is true for the overall probability of leaving Integration. Thus, the main patterns emphasized in Section 4 are already clearly visible in the raw data and remain essentially unchanged after accounting for persistent individual heterogeneity in the dynamic model.

The main discrepancies concern transitions out of Marginalisation. Here the model-based probabilities differ more visibly from the raw frequencies, particularly for national identity and media use. Since Marginalisation is relatively uncommon in the sample, these raw transition frequencies are based on substantially fewer observations and should be interpreted more cautiously.

The estimates in table 9 should be read as descriptive rather than structural or causal. They document the raw empirical transition patterns, but does not distinguish genuine state dependence from persistent individual heterogeneity, compositional differences, or differential selection into observed year-to-year intervals. For this reason, in Section 4 we estimate transition matrices using a correlated random-effects approach, which provides a more appropriate framework for assessing transition dynamics net of observed and unobserved individual differences.

B Structural binary system with correlated random effects.

To estimate the transition dynamics, we use a simultaneous dynamic binary-response model with correlated random effects, in the spirit of Honoré et al. (2023).

The two equations share an individual-specific intercept α_i, which captures persistent unobserved heterogeneity.

Following Wooldridge (2005), we specify the distribution of this intercept conditional on the individual's initial outcomes.

Conditional on the individual-specific intercept and the lagged outcomes $(y_{1,i,t-1}, y_{2,i,t-1})$, the joint distribution of the two current outcomes is represented by a four-cell binary-response system with a contemporaneous association parameter.

The linear indices are

$$\eta_{1,i,t} = \alpha_i - \frac{\kappa}{2} + \gamma_{11} y_{1,i,t-1} + \gamma_{12} y_{2,i,t-1}, \tag{B.1}$$

$$\eta_{2,i,t} = \alpha_i + \frac{\kappa}{2} + \gamma_{21} y_{1,i,t-1} + \gamma_{22} y_{2,i,t-1}, \tag{B.2}$$

where κ measures the relative propensity for y_2 versus y_1, and

$$\alpha_i \;=\; d_0 + d_1 y_{1,i,0} + d_2 y_{2,i,0} + d_3 (y_{1,i,0} y_{2,i,0}) + a_i, \qquad a_i \sim \mathcal{N}(0, \sigma^2). \tag{B.3}$$

Here $y_{j,i,0}$ are initial conditions (first observed values in the panel); (d_1, d_2, d_3) implement the Wooldridge initial-conditions correction.

Let ρ denote the contemporaneous association parameter. With (B.1), the four joint probabilities for $(y_{1,i,t}, y_{2,i,t}) \in \{(0,0), (1,0), (0,1), (1,1)\}$, conditional on a_i and the lags, are

$$D := 1 + e^{\eta_{1,i,t}} + e^{\eta_{2,i,t}} + e^{\eta_{1,i,t} + \eta_{2,i,t} + \rho},$$

$$\Pr(0,0 \mid a_i, \text{lags}) = \frac{1}{D}, \qquad \Pr(1,0 \mid a_i, \text{lags}) = \frac{e^{\eta_{1,i,t}}}{D},$$

$$\Pr(0,1 \mid a_i, \text{lags}) = \frac{e^{\eta_{2,i,t}}}{D}, \qquad \Pr(1,1 \mid a_i, \text{lags}) = \frac{e^{\eta_{1,i,t} + \eta_{2,i,t} + \rho}}{D}. \tag{B.4}$$

When $\rho = 0$, the two outcomes are conditionally independent given the individual-specific effect and lagged outcomes.

Likelihood, integration, and optimisation. Let $N_i(s \to d)$ denote the number of observed transitions from state s to state d for individual i.

A transition is included when both binary outcomes are observed in the current and immediately preceding survey records for the same individual.

Conditional on the residual random effect a_i, the individual likelihood contribution is

$$L_i(a_i) = \prod_{s=1}^{4} \prod_{d=1}^{4} \Big(\Pr(s \to d \mid a_i) \Big)^{N_i(s \to d)}, \tag{B.5}$$

where $\Pr(s \to d \mid a_i)$ is implied by (B.4) after setting the lagged pair corresponding to s. The unconditional log-likelihood integrates out a_i:

$$\ell(\psi) = \sum_i \log \int L_i(a_i)\, \phi\Big(\frac{a_i}{\sigma}\Big)\, \frac{da_i}{\sigma}, \tag{B.6}$$

with parameter vector

$$\psi = \big(\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}, \rho, \kappa, d_0, d_1, d_2, d_3, \ln \sigma\big), \tag{B.7}$$

and ϕ the standard normal density. The integral is approximated using 10-point Gauss–Hermite quadrature. Parameters are estimated by maximizing the log-likelihood using BFGS, initialized from smoothed empirical transition frequencies.

Recovering model-implied transition matrices. Given the estimated parameters $\hat{\psi}$, we recover the model-implied transition matrix.

Let

$$W_{is} = \sum_{d=1}^{4} N_i(s \to d)$$

denote the number of observed transitions originating from state s for individual i.

The individual-specific conditional mean of the random intercept is

$$\hat{m}_i = \hat{d}_0 + \hat{d}_1 y_{1,i,0} + \hat{d}_2 y_{2,i,0} + \hat{d}_3 y_{1,i,0} y_{2,i,0},$$

.

For each origin state s, we fix the lagged binary outcomes at the values corresponding to that state.

We then evaluate the four destination-state probabilities implied by Equation (B.4), integrating over the residual individual random effect using Gauss–Hermite quadrature.

Let $p_d(s, \alpha; \psi)$ denote the probability of destination state d when the previous state is s and the individual-specific intercept equals α.

The estimated transition probability is

$$\widehat{T}_{s,d} = \frac{\sum_{i \in \mathcal{I}} W_{is} \left[\frac{1}{\sqrt{\pi}} \sum_{k=1}^{K} \omega_k p_d \left(s, \hat{m}_i + \sqrt{2}\hat{\sigma} z_k; \hat{\psi} \right) \right]}{\sum_{i \in \mathcal{I}} W_{is}}, \tag{B.8}$$

where $\mathcal{I}$ denotes the estimation sample of first-generation immigrants, and z_k and ω_k are the Gauss–Hermite nodes and weights.

The expression is defined for origin states with at least one observed transition in the estimation sample.

For each origin state, individuals receive weights proportional to the number of observed transitions originating from that state.

Consequently, the estimated matrix represents an origin-state-exposure-weighted average of model-implied transition probabilities.

The distribution of initial conditions underlying this average may differ across origin states because the individual weights W_{is} are state-specific.

The residual random effect is integrated using its assumed distribution conditional on initial outcomes, rather than its posterior distribution conditional on the individual's complete observed history.

Stacking the four destination-probability vectors yields the 4×4 matrix $\widehat{T}$.

Each row sums to one, up to numerical precision.

Visualisation and frequency weighting. The Sankey diagrams provide a visual representation of the model-implied transition matrices.

To scale the links by the relative prevalence of each origin state, we multiply the corresponding row of the estimated matrix by its empirical initial-state share.

Formally, if $\hat{f}_s$ denotes the initial-state share, the plotted flow from state s to state d is

$$\text{flow}_{s \to d} = \hat{f}_s \widehat{T}_{sd}. \tag{B.9}$$

These flow masses combine the empirical initial-state distribution with the model-implied transition probabilities. They should not be interpreted as observed transition frequencies.

Interpretation. The dynamic coefficients $(\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22})$ capture state dependence across dimensions, allowing past commitmentin either dimension to influence current choices in both. The parameter κ shifts the relative baseline propensity toward the origin versus destination dimension, while ρ summarizes residual same-period association beyond the lags and random effects. The initial-conditions parameters allow the distribution of persistent individual heterogeneity to depend on initial identity states, under the maintained correlated random-effects specification. The resulting $\widehat{T}$ matrices, together with frequency-weighted Sankey plots, summarize the evolution of identity over time in a way that is both model-based and transparent for descriptive comparison.

C Correlates of Berry states and mobility

We complement the estimated transition probabilities with an exploratory analysis of the observable characteristics associated with the presence in Berry states and the transitios across them. These regressions address the descriptive question of what is the composition of states and which individuals experience transitions, rather than identifying the mechanisms of identity revision. They should therefore be interpreted as conditional associations, not as estimates of causal or structural effects.

C.1 Covariates and construction

Table 10 summarizes the covariates used in the exploratory regressions. Individual-level information comes mainly from the SOEP person-year files and covers demographic characteristics, migration history, economic resources, health, political orientation, family status, residence, religion, citizenship and settlement intentions, employment, and education. Within individuals, missing values are first filled forward and then backward using available observations for the same individual. For the small number of cases in which a covariate is never observed for an individual, the remaining missing values are replaced by the sample mean to preserve the estimation sample. This procedure is adopted to retain observations and should not be interpreted as a model of the missing-data process. Monetary variables are expressed in thousands of euros.

C.2 Compositional analysis of Berry states

Table 11 shows the composition of Berry states with respect to a selection of individual characteristics. The *All states* column provides a within-domain benchmark, computed over all person-year observations for which the Berry state in the relevant domain is observed.[11] Differences exist across both states and domains. In national identity,

[11] The domain-specific benchmarks in Table 11 need not coincide exactly with the descriptive means reported in Table 10. The latter are computed over the pooled person-year sample, whereas the former

Table 10: Descriptive statistics of regression covariates

Variable	Description	Mean	SD	Min	Max	Imputed
age	Age in years	42.966	14.386	17.000	91.000	No
age_arrival	Age at arrival in Germany	19.075	13.237	0.000	72.000	No
distrust_index_imp	Interpersonal distrust index	0.445	0.255	0.000	1.000	Yes
health_index_imp	Health index	0.819	0.287	0.000	1.000	Yes
income_imp	Individual income, thousand euros	1.915	1.796	0.000	50.000	Yes
wealth_individual	Individual net wealth, thousand euros	10.508	64.341	-173.500	1560.000	Yes
wealth_household	Household net wealth, thousand euros	23.399	109.751	-338.500	3429.000	Yes
left_right	Political orientation, 0 left to 10 right	4.613	1.420	0.000	10.000	Yes
gdp3_growth	Origin-country GDP per capita growth over previous 3 years	6.477	8.037	-24.400	30.951	No
female	Female	0.518	0.500	0.000	1.000	No
has_partner	Has partner	0.694	0.441	0.000	1.000	Yes
east_germany	Lives in East Germany	0.017	0.131	0.000	1.000	Yes
rural	Lives in rural area	0.203	0.399	0.000	1.000	Yes
relig_catholic_imp	Catholic	0.251	0.434	0.000	1.000	Yes
relig_protestant_imp	Protestant	0.145	0.352	0.000	1.000	Yes
relig_muslim_imp	Muslim	0.171	0.377	0.000	1.000	Yes
citizen_already_imp	Already a German citizen	0.000	0.000	0.000	0.000	Yes
citizen_intent_yes_imp	Intends to obtain German citizenship	0.160	0.367	0.000	1.000	Yes
citizen_intent_uncertain_imp	Uncertain about obtaining German citizenship	0.128	0.335	0.000	1.000	Yes
german_citizen_imp	German citizen	0.438	0.496	0.000	1.000	Yes
remain_germany_yes_imp	Intends to remain in Germany	0.765	0.424	0.000	1.000	Yes
emp_employed_imp	Employed	0.641	0.480	0.000	1.000	Yes
edu_high	High education	0.198	0.399	0.000	1.000	Yes
edu_medium	Medium education	0.533	0.499	0.000	1.000	Yes
self_emp	Self-employed	0.076	0.266	0.000	1.000	Yes

Notes: Descriptive statistics are computed at the person-year level. Monetary variables are expressed in thousands of euros. The "Imputed" column indicates whether missing observations are imputed in the data-processing procedure described in the text.

socioeconomic differences are relatively modest, while citizenship status stands out: 75.2% of assimilated person-year observations correspond to German citizens, compared with only 14.9% among the separated. Media use displays a clearer economic gradient, with separated individuals having lower average income than those in integration or assimilation (1.41 versus 2.05 and 2.01 thousand euros, respectively). The strongest socioeconomic differentiation appears in social interaction: integrated individuals are more often highly educated (21.7% versus 3.7% among the separated), employed (67.1% versus 49.8%), and have higher average income (1.99 versus 1.44 thousand euros). Political orientation, by contrast, varies little across states. Estimates for marginalisation should be interpreted cautiously because cell sizes are small, particularly for national identity.

C.3 Ever-mover profiles

We then ask whether observable characteristics are associated with experiencing at least one transition across Berry states during an individual's observed domain-specific history. For these regressions, person-year covariates are averaged over the years in which the relevant Berry state is observed for each individual. Table 12 suggests a few broad descriptive patterns. Religious affiliation is among the most recurrent correlates of cumulative mobility, with Catholic, Protestant, and Muslim indicators positively associated with ever-mover status across several domains. Employment is also positively associated with mobility in national identity and media use, whereas worse health is associated with a lower probability of moving in the media and social-interaction domains. In national identity, German citizenship and the intention to acquire citizenship are negatively associated with ever-mover status. This suggests greater stability among immigrants who pursue a formal commitment to the destination country. By contrast, the social-interaction domain displays a somewhat different profile, with mobility associated with age and age at arrival and negatively associated with income, female gender, and high education. Overall, the pattern is heterogeneous across domains rather than driven by a single common socioeconomic profile. Table 12 is a descriptive profile of

condition on the availability of the Berry-state measure in each domain.

Table 11: Composition of Berry states

Variable	All states	Int.	Assim.	Sep.	Marg.
		National identity			
years_germany	24.23	24.79	23.82	22.87	21.56
age	43.59	44.58	41.06	44.07	41.24
female	52.4%	53.2%	48.9%	55.5%	42.9%
edu_high	20.5%	21.6%	18.1%	20.5%	19.0%
emp_employed_imp	65.2%	65.8%	64.3%	65.2%	54.8%
income_imp	1.92	1.94	1.90	1.88	1.90
german_citizen_imp	43.3%	37.6%	75.2%	14.9%	19.0%
left_right	4.59	4.53	4.71	4.65	4.24
east_germany	1.6%	1.9%	0.5%	2.8%	0.0%
obs.	4,011	2,370	995	604	42
		Media use			
years_germany	24.16	25.20	25.07	18.26	21.76
age	43.16	45.47	41.87	47.21	39.63
female	51.6%	47.7%	51.8%	57.6%	54.0%
edu_high	20.6%	17.1%	22.9%	17.4%	17.0%
emp_employed_imp	65.5%	63.4%	68.7%	55.8%	59.4%
income_imp	1.93	2.05	2.01	1.41	1.65
german_citizen_imp	44.4%	27.5%	53.3%	34.3%	37.1%
left_right	4.59	4.75	4.55	4.57	4.35
east_germany	1.7%	0.9%	2.1%	2.3%	0.4%
obs.	4,023	920	2,436	443	224
		Social interaction			
years_germany	24.18	23.72	26.85	25.54	28.85
age	43.07	42.23	46.19	47.86	52.56
female	51.8%	52.9%	46.2%	46.5%	43.3%
edu_high	20.1%	21.7%	15.4%	3.7%	11.1%
emp_employed_imp	64.6%	67.1%	52.0%	49.8%	53.3%
income_imp	1.92	1.99	1.53	1.44	1.59
german_citizen_imp	44.8%	43.9%	69.2%	27.6%	28.9%
left_right	4.61	4.59	4.86	4.58	4.57
east_germany	1.7%	1.2%	5.5%	2.3%	5.6%
obs.	4,183	3,532	344	217	90

Notes: Means for continuous variables, percentages for binary variables. The table uses all person-year observations for which the domain-specific Berry state is observed. The *All states* column reports the corresponding mean or percentage within each domain.

cumulative mobility, and should not be interpreted as an estimation of the mechanisms studied in the main text.

Table 12: Correlates of ever-mover status

	National identity	Media use	Social network
age			0.003**
age_arrival			0.004**
health_index_imp		-0.205**	-0.122*
wealth_individual_thousand	-0.002*		
female			-0.051*
has_partner		0.085*	
rural		-0.088*	
relig_catholic_imp	0.098*	0.242***	0.177***
relig_protestant_imp	0.170***	0.117*	0.162***
relig_muslim_imp	0.208***	0.263***	0.097**
german_citizen_imp	-0.120**		
citizen_intent_yes_imp	-0.129**		
emp_employed_imp		0.116**	
edu_high			-0.101**
Observations	618	616	669
Movers	183	180	122
R^2	0.080	0.121	0.140

Notes: The dependent variable equals one if the individual experiences at least one transition across Berry states during the observed domain-specific history. For compactness, the table reports only covariates with at least one coefficient significant at the 10 percent level across the three domains; non-significant coefficients within reported rows are left blank.

References

Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. The quarterly journal of economics, 115(3):715–753.

Alba, R. D. and Nee, V. (2003). Remaking the American mainstream: Assimilation and contemporary immigration. Harvard University Press.

Baumeister, R. (1986). Identity: Cultural change and the struggle for self. Oxford University Press.

Berry, J. W. (1980). Acculturation as varieties of adaptation. Padilla, A., Ed., Acculturation: Theory, Models and Findings, Westview, Boulder, pages 9–25.

Bénabou, R. and Tirole, J. (2011). Identity, morals, and taboos: Beliefs as assets. The Quarterly Journal of Economics, 126(2):805–855.

Carillo, M. R., Venittelli, T., and Zazzaro, A. (2026). Immigrants' social identity, racial hate crimes and public backlash: Evidence from the "san gennaro massacre". European Economic Review, Forthcoming.

Chandra, K. (2012). What Is Ethnic Identity? A Minimalist Definition. In Constructivist Theories of Ethnic Politics. Oxford University Press.

Constant, A. F., Gataullina, L., and Zimmermann, K. F. (2009). Ethnosizing immigrants. Journal of Economic Behavior & Organization, 69(3):274–287.

Constant, A. F. and Zimmermann, K. F. (2008). Measuring ethnic identity and its impact on economic behavior. Journal of the European Economic Association, 6(2-3):424–433.

Herberg, W. (1955). Protestant–Catholic–Jew: An Essay in American Religious Sociology. University of Chicago Press.

Honoré, B. E., Hu, L., Kyriazidou, E., and Weidner, M. (2023). Simultaneity in binary outcome models with an application to employment for couples. Empirical Economics, 64(6):3197–3233.

Loomes, G. and Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. The economic journal, 92(368):805–824.

Loomes, G. and Sugden, R. (1986). Disappointment and dynamic consistency in choice under uncertainty. The Review of Economic Studies, 53(2):271–282.

Makabe, T. (1979). Ethnic identity scale and social mobility: the case of nisei in toronto*. Canadian Review of Sociology/Revue canadienne de sociologie, 16(2):136–146.

Minoru Masuda, G. H. M. and Meredith, G. M. (1970). Ethnic identity in three generations of japanese americans. The Journal of Social Psychology, 81(2):199–207. PMID: 5457351.

Phinney, J. S. (1990). Ethnic identity in adolescents and adults: review of research. Psychological bulletin, 108 3:499–514.

Portes, A. and Rumbaut, R. G. (2001). Legacies: The story of the immigrant second generation. Univ of California Press.

Runge, R. A. and Maehler, D. B. (2026). Navigating cultural identity: A 20-year longitudinal study of first-generation adult immigrants in germany. Journal of Personality and Social Psychology, 131(2):337.

Unger, J. B., Gallaher, P., Shakib, S., Ritt-Olson, A., Palmer, P. H., and Johnson, C. A. (2002). The ahimsa acculturation scale: A new measure of acculturation for adolescents in a multicultural society. The Journal of Early Adolescence, 22(3):225–251.

Wooldridge, J. M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. Journal of Applied Econometrics, 20(1):39–54.