Aug, 2026
GRINS DISCUSSION PAPER SERIES DP N° 127/2026 ISSN 3035-5576

GRINS

Dynamic Ethnosizing and dynamic identity

DP N° 127/2026

Authors:
Nicodemo De Vito, Fausto Galli, Giuseppe Russo, Ružica Savčić, Paolo Solimene

Aug, 2026
GRINS DISCUSSION PAPER SERIES DP N° 127/2026 **ISSN 3035-5576**

Dynamic Ethnosizing and dynamic identity

Nicodemo De Vito, Fausto Galli, Giuseppe Russo, Ružica Sav i , Paolo Solimene

KEYWORDS

Identity formation | immigration | acculturation | cultural transmission

JEL CODE

D91, F22, J15, K37, Z13

ACKNOWLEDGEMENTS

This study was funded by the European Union - NextGenerationEU, in the framework of the GRINS - Growing Resilient, INclusive and Sustainable project (GRINS PE00000018). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

CITE THIS WORK

Title: Dynamic Ethnosizing and dynamic identity. Publication Date: 2026.

Using rich panel data from the German SOEP (2010–2019), 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 combine a dynamic panel framework (System GMM) with a simultaneous two-equation dynamic binary model, allowing us to estimate both 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 identification with the destination culture weakens rather than strengthens.
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 behind identity revision and integration reversal.

DYNAMIC ETHNOSIZING AND DYNAMIC IDENTITY

Nicodemo De Vito (1,5), Fausto Galli (1), Giuseppe Russo*(1,2), Ružica Savčić (1,3), and Paolo Solimene (4,6)

1)DISES, University of Salerno.

2)CSEF and GLO

3)Department of Economics, University of Cyprus

4)Tinbergen Institute, University of Rotterdam

5)Bocconi University

6)DEAMS, University of Trieste

Abstract

Using rich panel data from the German SOEP (2010–2019), 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 combine dynamic panel framework (System GMM) with a simultaneous two-equation dynamic binary model, allowing us to estimate both 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 identification with the destination culture weakens rather than strengthens. 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 behind identity revision and integration reversal.

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

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

* This research is funded by the European Union - Next Generation EU, in the framework of the GRINS - Growing Resilient, Inclusive and Sustainable Project (GRINSPE00000018 – CUP UNISA D43C24002040006). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, and the European Union cannot be held responsible for them.

Keywords: Identity formation, immigration, acculturation, cultural transmission.
JEL classification: D91, F22, J15, K37, Z13.

Competing interests: The authors declare none.

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 marginalization (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) analyzes attachment 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 (2010–2019), we construct domain-specific Ethnosizer states (identity, media use, social interaction) and estimate their evolution with dynamic discrete-choice methods, in particular 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 marginalization, and to document when and how individuals move between them.

Second, on the *theoretical* side, we develop a dynamic model of identity choice with regret and rejoicing. In the model, integration is an investment that improves the distribution of future outcomes but entails an up-front cost. Realized outcomes are compared to counterfactuals, generating ex post regret or confirmation; these retrospective evaluations can trigger revisions of identity, including what we term *integra-*

tion reversal, where immigrants step back from previously integrated positions toward more separated ones.

Relative to Constant et al. (2009) and the subsequent Ethnosizer literature, 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 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. According to 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 the expectations and the realized returns to migration and integration can lead to regret the previous choices and revise the identity. By combining rich 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 attachments 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 reversions away from integration or assimilation. 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 deepening the beliefs.

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 integration reversal.

The paper proceeds as follows. Section 2 briefly reviews the essential literature on

identity, then it presents the construction of the Ethnosizer and the SOEP data. Section 3 shows GMM estimates of Ethnosizer score dynamics, and Section 3.2 develops and estimates the correlated random-effects bivariate dynamic logit model of transition. Section 4 introduces our identity-and-regret model and links its implications to the empirical transition patterns. Section 5 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 and 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, as important as it is, 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. *Changes* in identity and, in particular, its *reversibility* have been little investigated both in the empirical and theoretical literature. They are the focus of our contribution.

In section 5, we contribute to the development of a model in which immigrants' identity can be revised after the reconsideration of previous choices or unexpected shocks.
To explore the concept of identity among immigrants, there exist two main frameworks: 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 fully adapt to the host society without dismissing their original identity. Formally, this implies that the measures of commitment to the Origin and Destination country are statistically independent. 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.

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 split the plane into four quarters. Figure

(2) shows the "identity states" outlined by Berry (1980). In the top-left quarter, the individual is assimilated, showing a strong commitment to the host country and a low commitment to the origin country; in the top-right quarter the individual is integrated, showing commitment to both countries; in the bottom-right quarter the individual is separated, showing commitment only to the origin country; in the bottom-left quarter the individual is marginalised, showing no commitment to any country.
Intertemporal moves across the Berry states are the core of our analysis. Actually, our primary aim is finding empirical evidence that transitions between identity states are relevant. Then, we provide a theoretical foundation for these transitions.
One of the major results of our empirical analysis concerns "backward moves" 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 the Destination. Consequently, she ranks Berry's identity states as follows:

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

We say that integration reversal occurs if the identity in period 2 moves from a more preferred state to a less preferred state (e. g., from Assimilated to Integrated).

In our empirical strategy, we rely on two complementary econometric approaches. First, we estimate dynamic panel models of Ethnosizer scores using the Blundell–Bond System GMM estimator. This method is designed for large-N, small-T panels with endogenous regressors and allows us to capture the inertia and cross-influences among identity states while controlling for unobserved heterogeneity and potential endogeneity of lagged regressors. Second, we implement a model of simultaneous dynamic logits with correlated random effects to recover transition probabilities across the identity states defined by Berry (1980). This approach is particularly suited to studying the joint dynamics of destination and origin country attachments and delivers transition matrices that we visualize with Sankey diagrams. Together, the two methods provide a consistent picture. The System GMM results reveal strong persistence of identity states: assimilation and integration, in particular, are highly stable over time, while marginalisation shows weaker inertia and stronger connections to other states. The

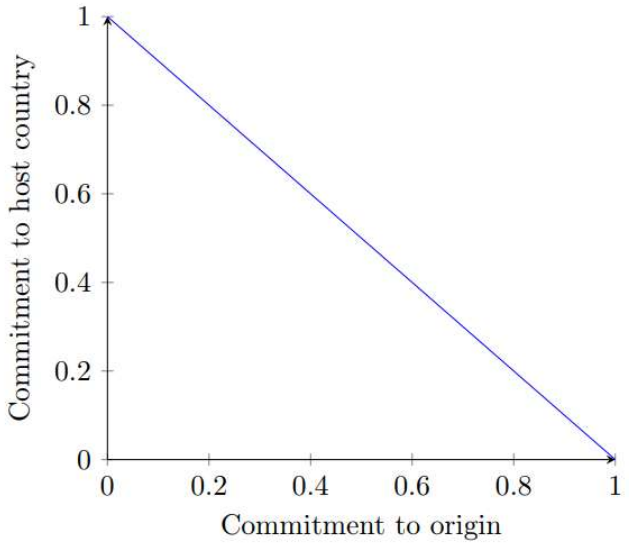


Figure 1: One-dimensional model

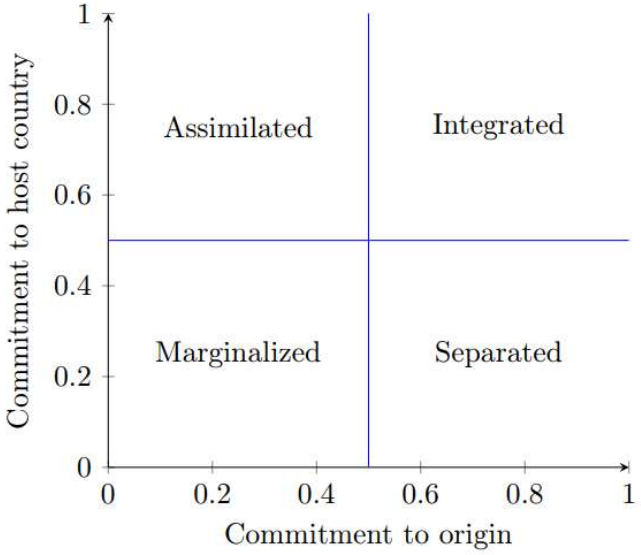


Figure 2: Two-dimensional model

logit framework confirms these patterns, highlighting the predominance of state persistence but also documenting non-negligible probabilities of moving across states. Both approaches underscore that identity formation is a sticky, path-dependent process, yet one that remains open to shifts.

2.1 The construction of the Ethnosizer

The Ethnosizer is a measure of identity introduced by Constant and Zimmermann (2008) and further developed by Constant et al. (2009), which translates Berry's (1980) bidimensional identity framework into an operational measure at the individual level. Specifically, they observe four domains — language, identity, media use, and social interaction — and classify each response according to the two relevant dimensions: attachment to the country of origin and attachment to the country of destination. Crossing these two dimensions yields the four identity states identified by Berry: assimilation, integration, separation, and marginalisation. They assign each individual a score from 0 to 4 in each quadrant, obtained by summing the contributions from the four domains. By construction, the quadrant scores add up to four, ensuring that the Ethnosizer-based measure captures the relative balance of states across all possible identity states (Constant and Zimmermann, 2008; Constant et al., 2009).
We exclude language from the Ethnosizer domains in our dynamic implementation. 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 domain 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 domains; namely, identity, media use, and social interaction. We work with SOEP-CORE waves 2010–2019; earlier years are dropped so all survey instruments are comparable.

Identity. Respondents rate how German they *feel* and how *attached* 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. Two frequency questions record whether the respondent *visited* or was *visited by* Germans (and foreigners) during the past year (1=*often*, 4=*never*). We average the two "German" items; a mean< 2 indicates frequent contact ($g = 1$). The same rule applied to the "Origin" items yields $o = 1$.

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 3x and carrying the common label set

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

2.2 Descriptive evidence

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
Marginalization score	.0839272	.2875961	0	2

In the descriptive evidence in table 1 integration emerges as the dominant state, with mean values around 2.3. Assimilation and marginalisation are relatively modest respectively with mean of 0.64 and 0.78, while separation is rare.

Table 1 reports summary statistics for the Ethnosizer identity scores. Integration emerges as the dominant state, with an average value above 2, indicating that most individuals display relatively high levels of simultaneous attachment to both origin and host societies.

Assimilation and marginalisation exhibit substantially lower mean values (around 0.6 and 0.8, respectively), suggesting that these states are less prevalent. Separation appears to be comparatively rare, with a mean close to zero and a limited dispersion.

Overall, the distribution of identity states is highly unbalanced, with integration clearly prevailing over the other configurations. This pattern is consistent with the idea that individuals tend to combine elements of origin and host cultures rather than adopting more polarized identity profiles.

3 Empirical methodology and results

3.1 Reduced-form GMM dynamic estimation of Ethnosizer score components

Table 2: System GMM estimates of Ethnosizer score components. Odd columns: no trend; even columns: quadratic time trend

	assimilated		integrated		separated		marginalised	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
L.assimilated	0.816***	0.750***	-0.0495	0.259	0.0665	-0.113	0.392	0.0106
	(0.226)	(0.154)	(1.340)	(0.234)	(0.139)	(0.0885)	(0.592)	(0.0284)
L.integrated	0.0542	0.298**	-0.0201	0.321	0.186	0.0806	0.193	0.0438
	(0.108)	(0.136)	(1.221)	(0.316)	(0.115)	(0.0647)	(0.491)	(0.0287)
L.separated	-0.180	-0.143	0.246	0.236	0.658***	0.692***	0.326	-0.0424
	(0.175)	(0.212)	(1.171)	(0.493)	(0.180)	(0.178)	(0.528)	(0.0474)
L.marginalised	0.286	-0.398	4.678	2.865	-0.0660	-0.300	1.079	0.617***
	(0.258)	(0.334)	(3.974)	(2.267)	(0.219)	(0.202)	(1.798)	(0.135)
t		-0.0656		0.182		0.0351		-0.0146
		(0.0489)		(0.118)		(0.0316)		(0.0162)
t^2		0.00430		-0.0107		-0.00296		0.000896
		(0.00356)		(0.00681)		(0.00238)		(0.00121)
Observations	3057	3057	3057	3057	3057	3057	3057	3057
Groups	550	550	550	550	550	550	550	550
Instruments	8	10	8	10	8	10	8	10
AR(1) p-value	0.005	0.061	0.559	0.872	0.092	0.026	0.747	0.008
AR(2) p-value	0.000	0.000	0.154	0.017	0.000	0.000	0.278	0.000
Hansen p-value	0.259	0.355	0.730	0.635	0.536	0.724	0.799	0.106

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2 presents dynamic panel-data estimates of the intertemporal evolution of the Berry identity states based on the Ethnosizer. The dependent variables in these regressions are the discrete score components—assimilation, integration, separation, and marginalisation. Each score records, for a given individual-year observation, how many of the three domains (identity, media use, and social interaction) fall into the corresponding Berry category. Thus, each score ranges from 0 to 3, and the four scores jointly summarize the composition of cultural orientations across domains.

Each score component is regressed on its own lag and on the lagged values of the other

three scores, allowing us to capture both persistence and cross-equation associations over time. The models are estimated using the Blundell–Bond System GMM estimator (Blundell and Bond, 1998), which is well suited to panels with large N and small T and in the presence of endogenous regressors. To ensure reliable inference, we employ collapsed instrument sets and restricted lag structures, thereby limiting instrument proliferation. Odd-numbered columns report specifications without time controls, while even-numbered columns include a quadratic time trend to capture common temporal dynamics.

These equations are best interpreted as reduced-form dynamic score equations. Their purpose is to summarize persistence and joint dynamics in the four score components [1]. The results point to substantial persistence in some score components, though with marked heterogeneity across dimensions.

Assimilation exhibits strong and statistically significant inertia, with autoregressive coefficients around 0.75–0.82. This indicates that assimilation, once established, tends to be highly stable over time. The inclusion of time controls slightly reduces the magnitude of the coefficient, suggesting that part of the persistence reflects common trends, but the overall picture of strong path dependence remains intact.

Separation emerges as similarly persistent. The lagged dependent variable is consistently large and highly significant, with coefficients around 0.65–0.70. This confirms that separation represents a stable and self-reinforcing component of the Ethnosizer score vector, largely driven by its own past values. The stability of these estimates across specifications further strengthens this interpretation.

By contrast, integration appears considerably less persistent. The autoregressive coefficient is not statistically significant and remains imprecisely estimated across specifications. Accordingly, the integration equation should be interpreted more cautiously than the assimilation and separation equations. In this reduced-form framework, the integration component appears noisier and less sharply identified.

Marginalisation displays an intermediate pattern. In specifications including time controls, the lagged dependent variable is positive and statistically significant, indicating

[1] We perform a structural analysis of transitions across identity states In the next subsection.

some degree of persistence. However, the magnitude and significance of the coefficient vary across models, pointing to a state that is less stable than assimilation or separation and potentially more sensitive to external conditions.

The interpretation of the cross-lag coefficients requires some caution because the four score components are jointly derived from the same three underlying domains. By construction, they are linked by a compositional constraint: increases in one component imply decreases in at least one other component. For this reason, the cross-lag terms are more naturally read as reduced-form cross-equation associations than as direct behavioral spillovers from one identity state to another. They remain informative about how the composition of the Ethnosizer evolves over time, and should not be overinterpreted as structural transition effects.

From this perspective, the broad pattern that emerges is one of strong own-component persistence and comparatively limited cross-component dependence. Assimilation appears largely insulated from the influence of the other components, evolving primarily through its own internal dynamics. Separation is also mainly driven by its own lag, with little evidence of systematic cross-component effects. Marginalisation exhibits somewhat more interaction with the other components, although these associations are not uniformly robust across specifications. The integration equation is the least stable, and its cross-lag terms are correspondingly harder to interpret sharply.

Specification tests broadly support the empirical strategy. The number of instruments remains modest relative to the number of groups, reducing concerns about overfitting. Hansen tests do not reject the validity of the instrument sets, while AR(1) tests behave as expected. The AR(2) results are more mixed and suggest that some equations are better behaved than others. Accordingly, the coefficients should not be interpreted as precise structural parameters; the more robust message is qualitative and concerns the heterogeneous degree of persistence across score components.

Overall, the findings highlight that the Berry identity states evolve gradually and are shaped by past positions, but with important differences across dimensions. Assimilation and separation emerge as highly persistent and self-reinforcing components, whereas integration is more fluid and less precisely estimated. Marginalisation occupies an intermediate position, combining some persistence with greater sensitivity to model

specification and external influences. Together, the GMM results provide a compact reduced-form picture of dynamic inertia in the Ethnosizer, while the interpretation of transitions across identity states is left to the structural transition model that follows.

3.2 Simultaneous dynamic binary-response model with correlated random effects

3.2.1 Methodology

We model the joint dynamics of two binary outcomes observed for each individual i across years t: $y_{1,i,t} \in \{0,1\}$ indicates engagement with the destination-country dimension (e.g. German-media use), and $y_{2,i,t} \in \{0,1\}$ indicates engagement with the origin dimension. This coding is aligned with the Ethnosizer construction in Section 2.1, where $(y_{1,i,t}, y_{2,i,t}) = (1,1)$ corresponds to integration, $(1,0)$ to assimilation, $(0,1)$ to separation, and $(0,0)$ to marginalisation.

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

$$s_{i,t} = 1 + y_{1,i,t} + 2y_{2,i,t} \quad \Longleftrightarrow \quad \begin{cases} s = 1: & (1,1) \text{ Integration}, \\ s = 2: & (1,0) \text{ Assimilation}, \\ s = 3: & (0,1) \text{ Separation}, \\ s = 4: & (0,0) \text{ Marginalisation}. \end{cases} \tag{3.1}$$

Our target is the 4×4 matrix T whose (s,d) entry is

$$T_{s,d} \;=\; \Pr\big(s_{i,t} = d \,\big|\, s_{i,t-1} = s\big), \tag{3.2}$$

i.e. the probability of moving from state s at $t-1$ to state d at t.

Structural binary system with correlated random effects. To estimate T we use a simultaneous dynamic binary-response model with correlated random effects in the spirit of Honoré et al. (2023). Let a_i denote an individual-specific intercept common to the two equations, capturing time-invariant heterogeneity and allowed to be correlated with the history via Wooldridge initial-conditions correction (Wooldridge, 2005).

Conditional on a_i and the lagged outcomes $(y_{1,i,t-1}, y_{2,i,t-1})$, the joint distribution of $(y_{1,i,t}, y_{2,i,t})$ is a joint four-cell binary system with an association parameter governing contemporaneous dependence. Write the linear indices

$$\eta_{1,i,t} = \alpha_i - \frac{\kappa}{2} + \gamma_{11} y_{1,i,t-1} + \gamma_{12} y_{2,i,t-1}, \tag{3.3}$$

$$\eta_{2,i,t} = \alpha_i + \frac{\kappa}{2} + \gamma_{21} y_{1,i,t-1} + \gamma_{22} y_{2,i,t-1}, \tag{3.4}$$

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

$$\alpha_i \;=\; d_0 + d_1 y_{1,i,0} + d_2 y_{2,i,0} + d_3 (y_{1,i,0} y_{2,i,0}) + a_i, \qquad a_i \sim \mathcal{N}(0, \sigma^2). \tag{3.5}$$

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 (3.3), 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

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

$$\begin{aligned}
\Pr(0,0 \mid a_i, \text{lags}) &= \frac{1}{D}, & \Pr(1,0 \mid a_i, \text{lags}) &= \frac{e^{\eta_{1,i,t}}}{D}, \\
\Pr(0,1 \mid a_i, \text{lags}) &= \frac{e^{\eta_{2,i,t}}}{D}, & \Pr(1,1 \mid a_i, \text{lags}) &= \frac{e^{\eta_{1,i,t} + \eta_{2,i,t} + \rho}}{D}.
\end{aligned} \tag{3.6}$$

This nests two dynamic logits when $\rho = 0$; $\rho \neq 0$ allows residual contemporaneous dependence.

Likelihood, integration, and optimisation. Let $N_i(s \to d)$ denote, for individual i, the count of transitions from state s at $t-1$ to d at t across the observed panel. Conditional on a_i, the individual contribution to the likelihood is

$$L_i(a_i) = \prod_{s=1}^{4} \prod_{d=1}^{4} \Big(\Pr(s \to d \mid a_i) \Big)^{N_i(s \to d)}, \tag{3.7}$$

where $\Pr(s \to d \mid a_i)$ is implied by (3.6) after setting the lagged pair corresponding to s. The unconditional log-likelihood integrates out a_i:

$$\ell(\psi) = \sum_i \log \int L_i(a_i)\, \phi\Big(\frac{a_i}{\sigma}\Big)\, \frac{da_i}{\sigma}, \tag{3.8}$$

with parameter vector

$$\psi = \big(\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}, \rho, \kappa, d_0, d_1, d_2, d_3, \ln \sigma\big), \tag{3.9}$$

and ϕ the standard normal density. The integral is approximated by Gauss–Hermite quadrature with K nodes; numerically, we evaluate the inner sum using a stable log-sum-exp transformation. Parameters are estimated by maximizing $\ell(\psi)$ via BFGS, with initial parameter values taken from reduced-form logits or smoothed empirical transition frequencies (with an optional short Nelder–Mead warm-up). Convergence tolerances on objective and parameters are set tightly.

Recovering model-implied transition matrices. Given $\hat{\psi}$ and for each origin state $s \in \{1, 2, 3, 4\}$, predicted destination probabilities are computed by fixing the lagged pair implied by s, evaluating the four joint probabilities for $(y_{1,t}, y_{2,t})$ at each Gauss–Hermite node, averaging with quadrature weights, and normalizing to one. Stacking the four destination vectors yields the 4×4 transition matrix $\widehat{T}$. Because the random effect enters additively and we integrate it out in prediction, the rows of $\widehat{T}$ represent population-average transition probabilities conditional on the previous state.

Visualisation and frequency weighting. Sankey diagrams display these transitions by mapping each row of $\widehat{T}$ into flows from state s to destinations d. To represent the prominence of each origin state in the sample at baseline, we multiply row s of $\widehat{T}$ by the empirical share of individuals whose first observed state is s . Formally, if $\widehat{f}_s$ is the initial-state share, the plotted flow mass from s to d is

$$\text{flow}_{s \to d} \;=\; \widehat{f}_s\, \widehat{T}_{sd}. \tag{3.10}$$

This preserves the transition structure while scaling links by the size of the origin pool, aiding visual comparison across domains.

Interpretation. The dynamic coefficients $(\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22})$ capture state dependence across dimensions, allowing past engagement in 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 controls $(y_{1,i,0}, y_{2,i,0}, y_{1,i,0}y_{2,i,0})$ absorb correlation between a_i and the process history, delivering a correlated random-effects estimator that is robust to nonrandom initial states. 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.

3.2.2 Results

This section presents the estimated transition matrices for the three domains of interest — media use, social contact, and 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). The four states correspond to: *Marginalisation* $(0,0)$, *Assimilation* $(1,0)$, *Separation* $(0,1)$, and *Integration* $(1,1)$. Rows sum to approximately one by construction. Sankey diagrams are used to visualize these estimated transitions in the main text, while the full numerical tables are reported in Appendix A.

Table 3: Summary of transition dynamics by domain

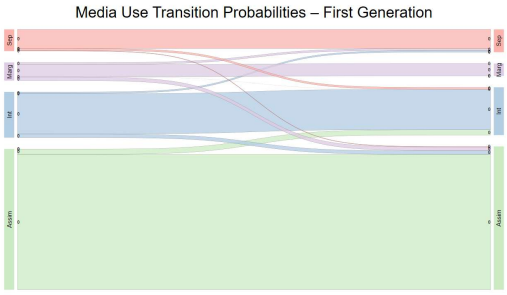
Domain	Mean persistence	P(int. from marg.)	P(int. from sep.)	P(out of int.)
Media use	0.842	0.003	0.089	0.116
Social contact	0.807	0.189	0.174	0.034
Identity	0.812	0.016	0.151	0.063

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 out of integration are defined as $1 - P(\text{integration} \to \text{integration})$.

Table 3 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 identity (around 0.01 or less), but much more frequent in social contact. Transitions from separation into integration are present in all domains, with estimated probabilities of 0.089 in media use, 0.174 in social contact, and 0.151 in identity. At the same time, the probability of leaving an integrated state is small but non-negligible—between 3% and 12% depending on the domain, providing a compact measure of exit from integration. 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 3, 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%) At the same time, a small but non-negligible fraction of integrated individuals move back into less integrated states in each period (roughly 10–12%), providing evidence of integration reversal in the media domain.

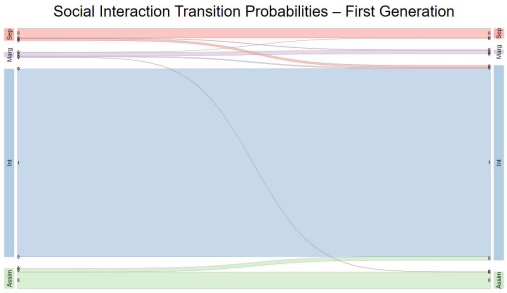


(a)

Figure 3: Estimated transition probabilities: media use

Social contact

Social contact also features strong state dependence, with mean persistence close to 0.80 (Table 3). 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. However, the probability of exiting an integrated state is very low (around 2–3%), indicating that once regular contact with both natives and co-ethnics is established it is rarely abandoned. Social interaction thus appears to be a domain where integration is both an important gateway state and a relatively safe one.



(a)

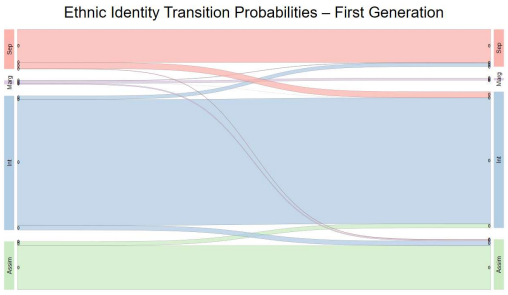
Figure 4: Estimated transition probabilities: social interaction

Identity

For 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 (around 15% in each case), indicating that "one-sided" identity profiles can evolve into more balanced ones over time. At the same time, a small share of integrated individuals revert to less integrated states in each period (around 6%), which we interpret as integration reversal. Together, the identity dynamics show that identity is both sticky and revisable.

3.2.3 Summary

Across all three domains, the estimated dynamics point to strong state persistence with gradual mobility. Integration and assimilation are highly stable states. For compact-



(a)

Figure 5: Estimated transition probabilities: identity

ness, only statistically significant transition probabilities are reported; omitted entries are not statistically different from zero at the 5% level.

4 A two-period model of identity formation

4.1 The model

In the previous sections we have shown how identity formation is a dynamic process, and, additionally, that identity is reversible. Here, we provide a novel theoretical model of identity formation that accounts for the observed transitions. 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 A, and $(1 - p_{i,t})$ to type α. Similarly, for each agent i and each period $t = 1, 2$, we let $q_{i,t}$ denote the probability assigned to type B, and $(1 - q_{i,t})$ to type β.

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

$$\begin{array}{lll} \text{belief on } (\alpha, \beta): & (1 - p_{i,t})(1 - q_{i,t}) & \text{(marginalized)} \\ \text{belief on } (A, \beta): & q_{i,t}(1 - p_{i,t}) & \text{(separated)} \\ \text{belief on } (\alpha, B): & p_{i,t}(1 - q_{i,t}) & \text{(assimilated)} \\ \text{belief on } (A, B): & p_{i,t}q_{i,t} & \text{(integrated)} \end{array}$$

This list provides a simple tool to represent identity states. We proceed as follows. First, we construct figure 6 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. $p_{i,t}$) $> 1/2$.[2] This divides the figure into four squares associated with the two levels of attachment 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 $(p_{i,t}, q_{i,t})$ falls into one square.[3]
Now, we define the prevalent identity as the the identity type that receives the highest probability mass. Formally,

$$\max\{p_{i,t}, (1 - p_{i,t})\} \cdot \max\{q_{i,t}, (1 - q_{i,t})\} > 1/4. \tag{4.1}$$

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

[2]Conversely, she is of type α if $q_{i,t} \leq 1/2$ and of type β if $p_{i,t} \leq 1/2$.
[3]Figure 6 can be interpreted as an Edgeworth box.

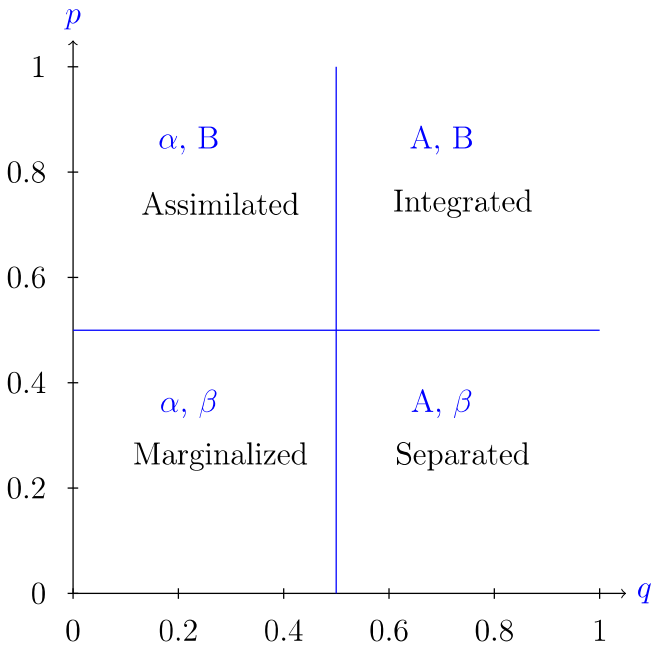


Figure 6: Cultural states in the (p, q) plane.

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.[4] 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

[4]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.

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.

4.2 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 $\theta_L < \theta_S < \theta_L$. In other words, investors who receive the low payoff (low net return) are worse off than non-investors[5] Investing is optimal when the following decision rule is satisfied:

$$a_i^1 = 1 \iff \mathbb{E}_{\mu_i}\left[u_i(1, .)\right] \geq \mathbb{E}_{\mu_i}\left[u_i(0, .)\right]. \tag{4.2}$$

[5]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.

where $u_i(.,.)$ is the VNM individual utility function, and $\mathbb{E}_{\mu_i}[u_i(.,.)]$ the expectation according to the probability measure μ_i. Equation 4.2 can be rewritten as

$$u_i(1,\theta_L)\mu_i(\theta_L) + u_i(1,\theta_H)\mu_i(\theta_H) \geq u_i(0,\theta_L)\mu_i(\theta_L) + u_i(0,\theta_H)\mu_i(\theta_H), \tag{4.3}$$

or, equivalently—via simple algebraic steps—as

$$\mu_i(\theta_H) \geq m\,\mu_i(\theta_L), \tag{4.4}$$

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 4.2 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.[6]

4.3 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.[7]

[6]Condition 4.2 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

[7]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).

4.4 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 $\theta = \theta_H$.

4.5 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.[8]
Residents in O receive a random revenue λ, which can be "low" or "high" —namely, $\lambda \in \Lambda = \{\lambda_H, \lambda_L\}$.
We order the outcomes in D and O as follows:

$$\lambda_L < \theta_L < \lambda_H < \theta_H. \tag{4.5}$$

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.[9] However, the point we wish to emphasize through the ordering (4.5) 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 revenue of the safe action in D, namely $\theta_L < \theta_H < \lambda_H$.

[8] Preliminary evidence from a survey we are collecting suggests that this regret dominates the regret about the integration investment.

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

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:

$$\lambda_L < \theta_L < \theta_S < \lambda_H < \theta_H. \tag{4.6}$$

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 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 $a_i^1 = 0, 1$.

4.5.1 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), $q_{i,2} \leq q_{i,1}$.

4.5.2 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, $q_{i,2} \leq q_{i,1}$.

4.5.3 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.

4.5.4 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.

4.6 Reconsideration and integration reversal

Section 5 provided a theoretical framework capable of explaining identity changes through a mechanism based on a retrospective consideration of one's past choices. More precisely, changes in p_{i2} and q_{i2} relative to their values in the first period can be interpreted as changes in identity. Note, however, that in the spirit of Berry (1980), identity changes only when moving from one of Berry's four states to another. This is evident from 2: as long as one remains within the same quadrant, their identity is barely affected. Intuitively, this means that immigrants for whom at least one of p_i and q_{it} is close to 0.5 can more easily experience a change of identity.
Our regret-based approach, togheter with the criteria introduced in Section 5.5, achieve the main objective of our model: providing theoretical foundations for the "integration reversal" defined earlier in Section 2.

5 Conclusions

We examined the dynamics of the Ethnosizer for immigrants in Germany across various measures of integration. Several major results emerge. First, identity is not static and can change over time. Second, while acculturation generally increases over time, the trend can reverse, suggesting the possibility of *integration reversal*. To obtain these results, we employed two complementary econometric strategies. The simultaneous dynamic binary-response model with correlated random effects complemented the GMM

analysis by modelling joint destination- and origin-country attachments and recovering transition probabilities across Ethnosizer states. Methodologically, the two empirical approaches played distinct roles: the System GMM estimates described persistence in the score-based Ethnosizer measures, while the simultaneous dynamic binary-response model with correlated random effects generated the transition matrices across Berry states. This dual approach provides both breadth and depth: it connects identity to observable characteristics while also uncovering its temporal path-dependence. The dynamic analysis reveals strong persistence across all states, with assimilation and integration displaying the highest stability. Separation and marginalization, though also persistent, appear more permeable to cross-influences from other states.

Overall, the empirical analysis shows that the identity states are considerably persistent, but the degree of stability varies across measures and domains. Transitions are not negligible, bringing to light the *reversibility* of the integration process. Although anecdotal evidence of this phenomenon exists, our analysis represents—to the best of our knowledge—the first attempt to identify it empirically and theoretically.

The reversibility of the integration process raises serious concerns in all societies where immigration is substantial. Usually, immigrants show the highest fertility rates in these societies. Their integration is key to the sustainability of the welfare systems and of 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. 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 confirm this intuition.

In addition, though 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 collecting, we think that retrospection is important to affirm identity. However, other mechanisms could also be relevant. We leave these avenues for future research.

Acknowledgements

We thank Pierpaolo Battigalli and Annamaria Menichini for their comments. We also thank participants to the Petralia Workshop in Applied Economics, the Final Spoke 3 Workshop in Naples, the 2026 EPCS Meeting in Madrid.

A Transition Matrices

Table 4: Estimated transition probabilities for media use.

State at $t-1$	(0,0)	(1,0)	(0,1)	(1,1)
(0,0)	0.684	0.199	0.114	0.003
(1,0)	0.000	0.939	0.000	0.037
(0,1)	0.000	0.027	0.861	0.089
(1,1)	0.000	0.081	0.035	0.884

Note: Only transition probabilities statistically significant at the 5% level are reported.
Statistically insignificant entries are set to zero.

Table 5: Estimated transition probabilities for social contact.

State at $t-1$	(0,0)	(1,0)	(0,1)	(1,1)
(0,0)	0.683	0.076	0.053	0.189
(1,0)	0.000	0.796	0.000	0.181
(0,1)	0.041	0.000	0.784	0.174
(1,1)	0.000	0.000	0.000	0.966

Note: Only transition probabilities statistically significant at the 5% level are reported. Statistically insignificant entries are set to zero.

Table 6: Estimated transition probabilities of identity states.

State at $t-1$	(0,0)	(1,0)	(0,1)	(1,1)
(0,0)	0.580	0.283	0.121	0.016
(1,0)	0.000	0.904	0.000	0.085
(0,1)	0.000	0.011	0.828	0.151
(1,1)	0.000	0.035	0.027	0.937

Note: Only transition probabilities statistically significant at the 5% level are reported. Statistically insignificant entries are set to zero.

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.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics, 87(1):115–143.

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

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.

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.