

Economic Models of Identity Formation and Cultural Transmission *

Alberto Bisin[†]

New York University, NBER, and CEPR

alberto.bisin@nyu.edu

Thierry Verdier[‡]

Paris School of Economics, ENPC-Institut Polytechnique de Paris, PUC-Rio and CEPR

thierry.verdier@psemail.eu

July 14, 2026

Abstract

In this paper we survey recent advances in the economic theory of identity formation. We restrict our analysis to models which conceptualize identity formation as embedded into an inter-generational cultural transmission process, and therefore have implications for the dynamics of social identity. Through the lens of a general framework linking identity formation to cultural evolution, we discuss issues related to the dynamics of oppositional, racial, or minority attitudes, cultural polarization, identity switching, and openness to alternative identities. An application of the set-up to multidimensional identities constructed from combinations of characteristics, provides conditions under which societies converge (or not) toward narrow identities centered on salient markers such as religion, ethnicity, or race. We conclude by outlining promising avenues for future research.

Journal of Economic Literature Classification Numbers: D1, J13, J15.

Keywords: Identity formation, Cultural transmission.

*This paper is prepared for the *Handbook of the Economics of Identity*; we thank the editors, George A. Akerlof and Rachel E. Kranton, for great support and comments.

[†]Department of Economics, New York University, 19 West Fourth Street, New York, NY, 10012, USA.

[‡]Paris School of Economics, 48 Boulevard Jourdan, 75014 Paris, France.

1 Introduction

Identity economics has grown rapidly over the past two decades. Drawing on cognitive and social psychology, it emphasizes how concerns about self-concept and social standing are important determinants of individual behavior. Along these lines, the theoretical literature on identity economics has taken three main distinct approaches to introduce notions of personal and social identity. The first, pioneered by [Akerlof and Kranton \[2000, 2005\]](#), incorporates social identity as utility payoffs tied to internalized prescriptions associated with group membership and social categorization.¹ In this class of models, individuals gain not only private benefits but also group-related benefits (e.g., status, self-esteem); they suffer utility losses when their behavior departs from group norms, when their traits diverge from group ideals, or when their personal characteristics are at variance with the ideal archetypes promoted within the group.² In many situations, identities reflect individual or collective choices: agents select which group to identify to, trading off the material and psychological costs and benefits of identification [[Shayo, 2009](#)].³

The second approach, associated with [Bénabou and Tirole \[2003, 2006, 2011\]](#), emphasizes instead cognition, assuming that individuals care about their self-image, social image, and deep values, but they lack full self-knowledge. Consequently, individuals manage beliefs and process information in ways that signal—both to themselves and others—commitment to valued identities.

Both these approaches typically take as fixed an underlying pattern of traits, categorizations, and psychological value sources that structure individual identity choices. Yet, some of these dimensions are culturally determined: they evolve slowly across generations, reshaping identity formation in the long run.

In this chapter we concentrate on a third approach which is centered on the study of the nature and implications of cultural evolutionary processes, offering an account of the dynamic evolution of social identity. Fundamentally, we conceptualize identity formation as embedded in the inter-generational cultural transmission process introduced by [Bisin and Verdier \[2001a\]](#),⁴ In this process, parents make deliberate socialization choices to shape their children’s preferences. These choices are however generally constrained by children’s role in forming their own cultural identity. Identity formation is then the result of a societal matching process which children generally act upon. Independently of the specifics of the identity formation mechanism, parental socialization choices will generally anticipate it and hence the socialization strategies they pursue will be accordingly affected⁵.

¹See also [Akerlof and Kranton \[2010\]](#) for a book-length treatment.

²This approach builds upon insights from the psychological literature on conformity [[Asch, 1955](#), [Cialdini and Goldstein, 2004](#)], norm formation [[Sherif, 1936](#)], self-categorization and ingroup-outgroup rivalry [[Tajfel, 1979](#), [Tajfel et al., 1979](#), [Tajfel, 1978](#), [Tajfel and Turner, 1979](#)].

³The preference-based modeling framework has been widely applied to many areas such as economic organizations [[Akerlof and Kranton, 2005](#)], labor markets [[Akerlof and Kranton, 2010](#)], education [[Akerlof and Kranton, 2002](#)], gender roles [[Bertrand et al., 2015](#)], distributive and identity politics [[Bonomi et al., 2021](#), [Ghiglino and Müller, 2026](#), [Ghiglino et al., 2021](#), [Lindqvist and Östling, 2013](#), [Shayo, 2009](#)], trade policy and attitudes towards globalization [[Grossman and Helpman \[2021\]](#), [[Holm, 2016](#)], and international integration [[Abramson and Shayo, 2022](#)], ethnic and civil conflicts [[Sambanis and Shayo, 2013](#), [Yuki, 2021](#)], war and nation building [[Sambanis et al., 2015](#), [Shayo, 2020](#)], rule-following norms [[Schneeberger and Krupka, 2021](#)], growth and development [[Skaperdas and Testa, 2025](#), [Yuki, 2023](#)].

⁴See [Boyd and Richerson \[1985\]](#), [Cavalli-Sforza and Feldman \[1981\]](#) for the study of cultural evolutionary processes with no parental socialization choice; see [Bisin and Verdier \[2025, 2010, 2023\]](#) for surveys of the economic approach and its place within the literature in the social sciences.

⁵[Doepeke and Zilibotti \[2017, 2019a,b\]](#) pursue important related work along these lines, studying parenting strategies which differ in whether and how they induce children merely comply with parental preferences (due to monitoring or fear of punishment) or instead internalize values—such as work ethic, patience, or respect for rules—which then guide behavior

Along these lines, after reviewing the structure of the canonical economic model of intergenerational cultural transmission - in Section 2 - we take a broad abstract view of identity formation within cultural evolution. In Section 3 we show how the rich insights from Identity economics [Akerlof and Kranton, 2000, 2010] can be incorporated into this class of models. We then delve into various applications in the literature. In Section 4 we survey models of identity formation which entail choices on the part of children which generally operate at the intensive margin - e.g., adopting oppositional attitudes within minoritarian cultures or racist attitudes within groups defined by given racial characteristics. In this context, we also introduce models studying cultural polarization as well as models of identity switching and cultural openness to different identities. In Section 5 we study identity formation in environments in which children choose identities constructed as a rich set of possible combinations of a vector of characteristics. In this context we break new grounds: the literature we survey does not embed the identity formation problem into cultural evolutionary dynamics, so we do it. We derive the general dynamical construct and apply it e.g., to characterize conditions under which identity formation converges - in the long-run - towards narrow identities - e.g., organized around high-salience markers like religion, ethnicity or race. Importantly, in Sections 4-5 we survey but furthermore extend existing work on identity formation - like Darity et al. [2006] on racialism, Shayo [2009] on general social identity, and Dasgupta and Goyal [2019] on multidimensional identities - to formally integrate it into dynamic models of cultural evolution. Section 6 concludes with remarks on future directions this literature may take.

2 Cultural Transmission and Identity Formation

Cultural evolution is the spread of cultural traits - interpreted as conveying socially relevant preferences, attitudes and beliefs - across individuals through social learning, socialization, teaching, or imitation. At the individual level, cultural evolution - by shaping preferences, attitudes and beliefs - affects behavior and decision-making. In this context, we conceptualize identity formation as embedded in the intergenerational cultural transmission process that drives cultural evolution. Specifically, cultural traits can be transmitted (vertically) from parents to children, and acquired (obliquely and horizontally) across and within generations among unrelated individuals through identity formation e.g., interacting with teachers and peers.

In this section we delineate the canonical model of cultural transmission, which operates through vertical transmission and identity formation [Bisin and Verdier, 2000, 2001b].⁶ While identity formation in this canonical model is represented by a simple societal mechanism, in the next sections of this chapter we will show how the model is enriched by allowing for various endogenous identity formation mechanisms.

2.1 The Canonical Cultural Transmission Model

Consider a population of a continuum of individuals each characterized by a binary cultural trait i . The population dynamics are assumed stationary and normalized to 1 for simplicity: reproduction is a-sexual and each parent has one child. Cultural transmission is the result of *direct vertical* parental socialization and identity formation in society at large. A parent with trait i successfully transmit vertically his/her

even in the absence of parental oversight. This internalization is essentially a mechanism of identity formation.

⁶The distinction between oblique and horizontal is immaterial in the continuous time approximation to cultural dynamics which the literature has largely adopted and hence we disregard it in this survey.

trait to his/her child with probability $0 \leq d^i \leq 1$. With probability $1 - d^i$ the child is not socialized and adopts a cultural identity through an oblique cultural transmission process where he/she matches at random with an individual from his/her parent's generation and acquires the trait of the match.

Formally, this cultural transmission mechanism is represented by transition probabilities, P^{ij} , encoding the probability that a parent with trait i has a child with trait j . Letting q^i denote the fraction of individuals with trait i in the population,

$$P^{ii}(d^i, q^i) = d^i + (1 - d^i)q^i \quad P^{ij}(d^i, q^i) = (1 - d^i)(1 - q^i), \quad i \neq j. \quad (1)$$

In the continuous time approximation, this process of cultural transmission results in the following cultural dynamics describing the diffusion of trait i in the population:

$$\dot{q}^i = q^i(1 - q^i)(d^i - d^j). \quad (2)$$

The economic approach to cultural transmission in [Bisin and Verdier \[2001b\]](#) allows parents to choose how intensively they directly socialize their children; that is, to choose socialization effort $0 \leq d^i \leq 1$, at some cost $c(d^i)$.⁷ Parents are motivated to cultural transmission by paternalistic preferences over the traits that their children can acquire. Consider first the simplest case in which a parent with trait i obtains an exogenous payoff V^{ij} if his/her child acquires trait j , assuming parents prefer children with the same trait as theirs: $V^{ii} > V^{ij}$ whenever $i \neq j$. A parent with trait $i \in \{a, b\}$ will then solve the following socialization problem:

$$\max_{d^i} P^{ii}(d^i, q^i)V^{ii} + P^{ij}(d^i, q^i)V^{ij} - c(d^i); \quad (3)$$

where $P^{ii}(d^i, q^i)$ and $P^{ij}(d^i, q^i)$ are given by (1).

Denote $\Delta V^i = V^{ii} - V^{ij}$ the parental preferences for socialization to cultural trait i , which the literature also refers to as *cultural intolerance* of cultural group i . The first-order conditions for i type are:

$$(1 - q^i)\Delta V^i = c'(d^i). \quad (4)$$

It immediately follows that socialization effort d^i increases with parental socialization preferences ΔV^i . Most importantly, in this specific formalization of the cultural transmission process, d^i decreases with the relative population size of cultural group i , q^i . Indeed, parents have less incentives to socialize their children when they belong to a cultural majority in the population as children tend to adopt the cultural identity of the majority in society. We say that, in this case direct vertical transmission act as a *cultural substitute* to identity formation in the cultural transmission mechanism.

It then follows that - under *cultural substitution* - the cultural transmission process tends to keep the cultural dynamics away from the boundaries $q^i = 0$ and $q^i = 1$ and an *heterogeneous* cultural distribution $0 < q^i < 1$ emerges as a stationary state from all interior initial condition of the population dynamics.⁸

$$q^i = \frac{\Delta V^i}{\Delta V^i + \Delta V^j}. \quad (5)$$

⁷The cost function $c(d)$ is assumed to be an increasing convex function with the Inada conditions: $c(0) = c'(0) = 0$, and $\lim_{d \rightarrow 1} c(d) = \lim_{d \rightarrow 1} c'(d) = +\infty$.

⁸This is in contrast with the case in which each d^i is exogenous, an assumption generally adopted in cultural evolution studies, starting from [\[Boyd and Richerson, 1985, Cavalli-Sforza and Feldman, 1981\]](#). In this case, in fact, the dynamics drive towards cultural homogenization with a *monomorphic equilibrium* (with $q^i = 1$ if $d^i > d^j$ or $q^i = 0$ if $d^i < d^j$). Equation (2) is actually a simple version of the replicator dynamics in evolutionary biology for a two-trait population dynamic model, where $(d^i - d^j)$ can be interpreted as the relative cultural fitness of trait i compared to trait j .

3 Identity Formation

Although parents make deliberate socialization choices to shape their children's preferences, vertical transmission is generally constrained by children's own active role in selecting preferences and cultural traits, that is, in forming their own cultural identity.

In the canonical model of cultural transmission identity formation is modeled as the result of a societal matching process which abstracts from any choice on the part of children. Rich insights from the economics of identity formation in [Akerlof and Kranton, 2000, 2005] can however be easily incorporated into the model. We distinguish broadly between models in which identity formation is modeled as a societal mechanisms and models which involve an explicit direct choice on the part of children.⁹ Independently of whether the socialization process involves direct identity adoption, parental vertical transmission choices will anticipate it and socialization patterns will be accordingly affected.

3.1 Societal Mechanisms of Identity Formation

The societal matching mechanism of identity formation which characterizes the canonical model of cultural transmission in the previous section gives rise, as we noted, to *cultural substitution* and hence induces an heterogeneous distribution of the population by cultural trait at the stationary state. Consider instead the case in which identity formation on the part of children requires psychological reinforcement mechanisms whereby the successful transmission of a particular cultural type may be only reached after a child has been exposed to a threshold number of similar role models, e.g., teachers, role models, peers. In this case identity formation may introduce forms of cultural complementarity.

As an illustration, consider the following societal identity formation mechanism studied by Bisin and Verdier [2001b]: direct parental socialization effort only pre-disposes a child to the parent's trait - which is adopted only if the child is also randomly matched with another individual of the same trait; otherwise the child is matched again with an individual in the population and adopts his/her trait. In this case

$$P^{ii}(d^i, q^i) = d^i q^i + q^i \quad P^{ij}(d^i, q^i) = (1 - d^i q^i) + 1 - q^i.$$

It is straightforward to see that a dimension of *cultural complementarity* is introduced in the cultural transmission process. It is not the case that d^i decreases with q^i and, as a consequence, *culturally homogeneous* stationary states - with $q^i = 1$ or $q^i = 0$ - arise from the dynamical system.

An important dimension of models of identity formation regards the study of how parents interact with the children adoption mechanisms, whether e.g., they seek to mitigate or manage them by limiting their children's social interactions. Consider an environment where parents have the capacity to affect their children social matching process by biasing the set of contacts they are exposed to when not directly socialized. More specifically, consider that a child's probability of matching with an individual with trait i is denoted $Q^i(s^i, q^i)$, where $0 \leq s^i \leq 1$ represents a choice variable of the parent affecting the cultural segregation of the environment the child is exposed to. Abstracting from specific details of the matching process, parental socialization is in this context multidimensional, in that it involves different margins of

⁹We refer to children as the parents' offspring in the model - even though they might be adults when engaging in identity formation.

transmission: direct socialization effort d^i and cultural segregation s^i .¹⁰

Given the contact process $Q^i(s^i, q^i)$, assume that the identity formation process of the child can be encapsulated into an adoption rule that generates a probability $T^i(s^i, q^i)$ that the child adopts trait i .¹¹

The optimal direct socialization rates d^i and segregation bias s^i are then the solution of the following problem:

$$\max_{(d^i, s^i)} P^{ii}(d^i, s^i, q^i) V^{ii} + \left(1 - P^{ii}(d^i, s^i, q^i)\right) V^{ij} - c(d^i, s^i);$$

where $P^{ii}(d^i, s^i, q^i) = d^i + (1 - d^i)T^i(s^i, q^i)$ and $c(d^i, s^i)$ is the parental socialization cost. The identity adoption mechanism operating through $T^i(s^i, q^i)$ affects the cultural dynamics through parental choices of d^i and s^i . Assuming that socialization costs satisfy some standard regularity assumptions, it can be shown that when direct socialization and segregation bias appear as complements in the socialization technology - that is, when $\frac{\partial^2 C^i}{\partial d^i \partial s^i} < 0$ - then *cultural substitution* prevails. Namely the optimal socialization choices d^i and s^i are decreasing in the frequency of the parent trait q^i . It follows that *cultural heterogeneous* obtains as a stationary state of the dynamics.

3.2 Identity Choice and Social Interactions

In the canonical model of cultural transmission - as well in the models in the previous section in which identity formation is the result of a matching process affecting children's social interactions - parental socialization preferences ΔV^i are exogenous. But in models which involve an explicit identity choice on the part of children, this choice will generally depend on the socio-economic environment the children face, which their parent will rationally anticipate. Consequently, parental socialization preferences will be explicitly derived from indirect utilities determined at the socio-economic equilibrium of their children and hence they will generally be frequency-dependent; that is, they will have the form $\Delta V^i(q^i)$, depending on q^i . Generally, in this literature, parental preferences for socialization are constructed under a formulation of imperfect altruism referred to as *imperfect empathy* [Bisin and Verdier, 2001b]: parents evaluate their children's actions with their own (the parents') utility.¹²

¹⁰A simple example may take the following form :

$$Q^i = Q^i(s^i, q^i) = s^i + (1 - s^i)q^i;$$

where $s^i \in [0, 1]$ represents the segregation bias chosen by the parent. Typically when $s^i = 0$, there is random matching and $Q^i(0, q^i) = q^i$. When instead $s^i > 0$ there is a positive bias to match towards the trait of the parent i . Indeed when $s^i = 1$, identity formation is completely biased towards the cultural variant of the parent and the child is only exposed to his/her specific cultural type.

¹¹For instance, under an adoption rule specifying that an (odd) number N of matches is randomly drawn and the cultural trait of the majority of these is adopted:

$$T^i(s^i, q^i) = \left(\frac{N}{\frac{N+1}{2}} \right) \left(Q^i(s^i, q^i) \right)^{\frac{N+1}{2}} \left(1 - Q^i(s^i, q^i) \right)^{N - \frac{N+1}{2}}.$$

Alternatively, for an extreme conformity mechanism, specifying that cultural type i is adopted only if all the matches drawn are of the same type i :

$$T^i(s^i, q^i) = \left(Q^i(s^i, q^i) \right)^N.$$

¹²While biasing parental preferences, imperfect empathy does not necessarily imply that in equilibrium parents prefer children sharing their own cultural trait type.

The implications of the endogeneity of ΔV^i for socialization and population dynamics need to be derived case-by-case. A reduced form analysis is however useful to clarify what to look for in the applications.¹³

Consider parental socialization preferences of the general form $\Delta V^i(q^i)$. The parental socialization problem then determines $d^i(q^i, \Delta V^i(q^i))$ so that:

$$\frac{d}{dq_t^i} d^i(q^i, \Delta V^i(q^i)) = \underbrace{\frac{\partial d^i}{\partial q_t^i}}_{\text{direct effect}} + \underbrace{\frac{\partial d^i}{\partial \Delta V^i} \frac{d\Delta V^i}{dq_{t+1}^i}}_{\text{feedback effect}} \underbrace{\frac{dq_{t+1}^i}{\partial q_t^i}}_{\text{expectations effect}}.$$

Under cultural substitution, $\frac{\partial d^i}{\partial q_t^i} < 0$. But the feedback effect $\frac{\partial d^i}{\partial \Delta V^i} \frac{d\Delta V^i}{dq_{t+1}^i}$ can be positive or negative: an anticipated larger frequency of trait i in the children's generation, may increase or decrease the parental socialization preferences.

In the continuous time approximation, the dynamical system for the evolution of cultural traits - when paternalistic socialization preferences are endogenous - can be written as:

$$\dot{q}^i = q^i(1 - q^i) [d^i(q^i, \Delta V^i(q^i)) - d^j(1 - q^i, \Delta V^j(1 - q^i))].$$

In this case, a sufficient set of conditions to obtain a heterogeneous stationary state cultural distribution requires cultural substitution as well as an additional condition on $\Delta V^i(q^i)$ - called *strategic substitution* in (Bisin and Verdier [2001b]):

$$\frac{d}{dq_i} \Delta V^i(q^i) < 0.$$

In words, *strategic substitution* implies that cultural minorities enjoy relatively larger gains from socialization, regardless of the specific socialization channel. It follows that if the social environment exhibits strategic substitution and, in addition, cultural substitution, then from any interior initial state, the cultural dynamics converge to a heterogeneous stationary state $0 < q^i < 1$.

By contrast, under strategic complementarity - that is, if $\frac{\partial}{\partial q_i} \Delta V^i(q^i) > 0$ - minorities experience smaller, and potentially even negative, socialization gains. In that case, whether heterogeneous stationary state is sustained in the limit depends on the strength of the cultural substitution effect. The precise way in which $\Delta V^i(q_i)$ depends on frequency is context-specific. In the following section we will map different identification formation models of interest into frequency dependent parental socialization preferences $\Delta V^i(q_i)$.

4 Social Identities at the Intensive Margin

In this section we study models where identity formation, while embedded in the intergenerational cultural transmission process, entails explicit choices on the part of children, after the vertical transmission effort of their parents. In particular, we study various specific social identities of interest in the social sciences, studied in the literature. These are environments where children's choices affect the intensity of their attachment to the cultural trait transmitted by parental socialization. In the next section we study models where the choice set entails richer opportunities, that is, distinct social groups, at the extensive margin, defined by a broad set of characteristics.

¹³We follow Bisin and Verdier [2001b, 2010] strictly here.

4.1 Oppositional Identity

The sociological literature studying social identity emphasizes two distinct - opposed in fact - motivation driving identity formation: *cultural conformity* and *cultural distinction*. In the first case group identity is driven by preferences for inclusiveness: the more common a trait is in the population, the more individuals want to express it; see for instance, *assimilation theories*, in political science and sociology (Gordon [1964]; Moghaddam and Solliday [1991]), or *contact theory* in social psychology (Allport [1953]). The alternative view suggests that group identity is motivated by preferences to keep a distinctive cultural heritage which generates a sense of positive *distinctiveness*; see theories of *multiculturalism* (Glazer and Moynihan [1970]) and *social identity theory* in social psychology (Tajfel et al. [1981]; Turner [1982]; and Abrams and Hogg [1988]).¹⁴

An interesting illustration of the roles of *cultural conformity* and *cultural distinction* in identity formation mechanisms is the study of the formation of oppositional identities in minority communities by Bisin et al. [2011]. Oppositional identity on the part of specific social groups occur when individuals choose cultural categorizations and prescriptions defined in opposition to the categorizations and prescriptions of the mainstream group [Battu et al., 2007]. The concept of oppositional identity maps onto several well-documented real-world cases. First of all, Fordham and Ogbu's classic 1986 study [Fordham and Ogbu, 1986] documenting - by means of ethnographic data from a high school in Washington, D.C., that Black students were intentionally underachieving to avoid being accused of acting White by their peers, framing this as a defensive identity response to a history of exclusion from the rewards of achievement.¹⁵ Secondly, oppositional identity manifests itself as reactive ethnicity among second-generation immigrants, as ethnographically documented by [Portes and Rumbaut, 2001], e.g., in the case of Mexican adolescents reacting to a California ballot measure that would have denied undocumented immigrants public services in 1994; or in the case of Haitian immigrant families holding strongly their own ethnic identity in opposition to a stigmatized generic "Black American" identity. More formally, reactive ethnicity has also been statistically documented by Kalmijn and Kraaykamp [2018] for second-generation Moroccan and Turkish migrants in the Netherlands.

To illustrate how oppositional identity formation may be shaped by the interaction of the opposing forces of cultural conformity and cultural distinction, consider a set-up in which children inherit or adopt either a mainstream or oppositional attitude, but oppositional individuals can also make an extra effort to reaffirm that identity. Parents anticipate these identity choices when deciding how strongly to transmit their own cultural attitude. As a result, cultural transmission depends on how common each identity already is in the community. If conformity dominates, oppositional identity can become self-reinforcing and may lead to multiple long-run cultural outcomes; if distinction dominates, the society tends toward a single mixed outcome where both identities persist. Typically oppositional identity becomes more prevalent when reaffirming it is more rewarding or less costly.

More formally, consider a minoritarian cultural group interacting with the rest of society. Individuals in the group display two possible cultural attitudes/traits: oppositional and mainstream - respectively $i = o, m$. Children i) choose an abstract socio-economic action x that depends on the nature of their cultural attitudes i ; and ii) interact socially within their cultural group. Individuals with the oppositional

¹⁴Note that *Cultural distinction*, as defined here, is a property of individual preferences. It is related but distinct from *cultural substitution*, which is a property of socialization mechanisms.

¹⁵See also Austen-Smith and Fryer Jr [2005].

trait have an opportunity to reaffirm their identity by taking an identitarian effort v . Such effort increases the benefit of matching with another oppositional individual and also reduces the cost of matching with an individual of the mainstream type. Mainstream individuals do not have the choice of reaffirming their identity. Formally, preferences by attitude type are as follows:

$$\begin{aligned} V^o(x, v) &= u^o(x) + q^o [B + \beta v] - (1 - q^o) [c - \gamma v] - j(v) \\ V^m(x) &= u^m(x) + q^o B - (1 - q^o) C; \end{aligned}$$

where q^o is the fraction of oppositional individuals in the community, B is the benefit of interacting with individuals of the same cultural attitude, c is the cost of interacting with individuals of the other type; furthermore, the identitarian effort v at a cost $j(v)$ allows an oppositional individual to increase his/her utility from interactions with individuals sharing the same attitude by βv and to reduce the cost of interactions with individuals carrying the opposite attitude by γv (with β and γ positive).

Once socialized to a cultural attitude i , a child chooses a socio-economic action x and, if oppositional, an identitarian effort v . The socio-economic action of an individual i is simply $x^i = \arg \max u^i(x)$, for $i = o, m$. Under straightforward regularity conditions, the optimal identitarian action $v(q^o)$ is increasing or decreasing in q^o depending on whether β is greater or smaller than γ . The first case represent preferences characterized by *cultural conformism*: the more common the trait, the more you want to affirm it. On the opposite, $\beta < \gamma$ represents *cultural distinction*: the less common the trait, the more the individual wants to reaffirm it.

In this environment, since the children's identity choices depend on the distribution of cultural attitudes in the population q^o - and since parents anticipate these choices - paternal motivations for cultural transmission are frequency dependent, that is, take the form $\Delta V^o(q^o)$ and $\Delta V^m(q^o)$. This dependence reflects the fact that identity choices occur in a context of social interactions where different cultural types interact and respond differentially.¹⁶ Specifically, the paternal motivations for cultural transmission have the following form:

$$\begin{aligned} \Delta V^o(q^o) &= \overline{\Delta V}^o + [q^o \beta + (1 - q^o) \gamma] v(q^o) - j(v(q^o)) \\ \Delta V^m(q^o) &= \overline{\Delta V}^m + j(v(q^o)). \end{aligned} \tag{6}$$

They have two components. The terms $\overline{\Delta V}^o = u^o(x^o) - u^o(x^m)$, and $\overline{\Delta V}^m = u^m(x^m) - u^m(x^o)$ are the direct parental socialization preferences deriving from *imperfect empathy*. The term $j(v(q^o))$ in the preferences for socialization of mainstream parents reflects the fact that, again by imperfect empathy, the cost of re-affirming an oppositional identity is a useless activity when evaluated with their utility (of mainstream parents).¹⁷

The dynamics of cultural attitudes in this society, therefore, are a special case of the ones delineated in Section 3.2. They depend on whether the frequency-dependence induces strategic complementarity or strategic substitutability in transmission process. Interestingly, we can show that, in the context we are analyzing, strategic complementarity and substitutability are directly mapped onto the distinction between

¹⁶Indeed, from (6), the frequency dependence on q^o disappears when children identity choice v is restricted to be $v = 0$. In such a case, parents socialize exactly according to the direct paternalistic motivations $\overline{\Delta V}^o$, and $\overline{\Delta V}^m$.

¹⁷It is worth noting that, if parents had full control on the identitarian action v , they would impose their preferred identity action ex-ante, whatever the cultural type taken by the children after socialization: $v = v^o(q)$ for an oppositional parent and $v = 0$ for a mainstream parent. Consequently parental socialization effort would coincide with those arising in a society with no identity formation choice.

cultural conformism and cultural distinction. Specifically, under cultural conformism - that is, $\beta > \gamma$ - identity formation induces strategic complementarity in the transmission of the oppositional trait o and strategic substitutability in the transmission of the mainstream trait m . The opposite occurs under cultural distinction, when $\beta < \gamma$.¹⁸

Under regularity conditions, we can show that under cultural conformism multiple heterogeneous cultural stationary states may obtain in the long-run. Conversely under cultural distinction, the dynamics converges to a unique heterogeneous cultural stationary state q^i such that $\frac{q^o}{1-q^o} = \frac{\Delta V^o(q^o)}{\Delta V^m(q^o)}$. In both cases, simple and intuitive comparative statics can be obtained for the stable long run cultural steady state(s): an increase β or in γ leads to a larger long run fraction of oppositional people in the community. Similarly, a downward shift of the cost $j(v)$ affects positively the long run fraction of individuals with oppositional traits.¹⁹

4.2 Racist Identity

A large literature in sociology and in political science conceptualizes identity formation as a fundamental determinant the cultural dynamics of racial identity norms. The fundamental tenet of this view - sometimes referred to as *social constructionism* - is that race and racial categories are not fixed, natural, or biological facts but are continuously produced and contested through political, social, and historical processes: they are drawn, contested, and redrawn, e.g., by immigration policy, social movements, and social interaction [Cornell and Hartmann, 2006, Omi and Winant, 2014].²⁰

As a consequence, in empirical work in developmental psychology the dominant older paradigm — measuring ethnic-racial identity (ERI) once and treating it as a stable trait — is giving way to intensive longitudinal designs. Gonçalves et al. [2025] used repeated within-person measurements of Black adolescents in Chicago and found that fluctuations in positive ethnic-racial affect track discrimination-based stress in ways that differ substantially across individuals.²¹ Furthermore, along these lines, quasi-experimental designs test causal claims about identity formation; see e.g., Gillespie et al. [2025]. More generally, economic history has produced the sharpest evidence that racial identity responds to incentives rather than being fixed: for instance, Dahis et al. [2019] linked U.S. census records from 1880–1940 and documented large numbers of Black men reclassifying as white over time; Bodenhorn and Ruebeck [2003] made a similar argument for the antebellum period, modeling whether free mixed-race individuals adopted a distinct “mulatto” identity as a strategic response to the pecuniary and psychic costs whites imposed on that category.

A seminal theoretical paper underlying this literature is Darity Jr et al. [2006]. It focuses on the per-

¹⁸In an empirical application of the framework - using data from the Fourth National Survey of Ethnic Minorities in the United Kingdom, Bisin et al. [2006] find evidence that intense ethnic and religious identities are generally formed as cultural distinction mechanisms.

¹⁹More generally, the analysis in Bisin et al. [2011] allows for oppositional attitudes of the majority group - e.g., as form of racism - which might increase the incentives of individuals in the minority to adopt oppositional identities. These complementarities might induce social multiplier effects and/or multiple social steady state equilibria.

²⁰The origin of this view can be traced to W.E.B. Du Bois who, in *The Souls of Black Folk* [Du Bois, 1903], points to the idea that Black Americans experience identity through the eyes of a hostile majority, always measuring themselves by a tape that despises them.

²¹See also Fryer et al. [2012], who studies mixed-race (Black-white) adolescents using survey data and documents that they fare worse than either monoracial Black or white peers on a range of social and behavioral outcomes — risky behavior, social integration, academic engagement — even after controlling for socioeconomic background.

sistence of race as a socially constructed identity, rather than a phenotypical character along these lines. The authors construct a simple evolutionary game set-up of a society composed of two fixed racial or ethnic groups. While group membership itself is not chosen, individuals can choose how they interpret and express that membership: they may be *individualists*, who behave in a race-neutral way, or *racialists*, who attach strong identity significance to their group. Social interaction matters because people are randomly matched either within or across groups, with the parameter of social segmentation determining how often cross-group encounters occur. Racist identity has two sides: it gives psychological or expressive benefits when interacting with in-group racialists, but it can also serve as a defensive posture when facing racist behavior from the other group. Thus, racism is not simply a taste; it is partly sustained by the social environment in which people expect either solidarity from insiders or hostility from outsiders.

These different elements can be easily embedded into an intergenerational cultural transmission framework in which these interaction payoffs can be directly connected to parents' decisions to socialize children. In such a set-up, parents transmit either an individualist or racist identity within their own group, but their incentives depend on the current prevalence of racism in both groups. If many people in the other group are racist, parents have weaker incentives to transmit individualism, because a race-neutral child may be more exposed to hostile or exclusionary out-group behavior. At the same time, they have stronger incentives to transmit racism, because racist identity becomes more useful as a protective or affirming response. Similarly, if many people in one's own group are racist, racist identity becomes more attractive because it generates stronger in-group recognition and emotional payoff.

The key intuition is therefore one of "cross-group reinforcement": racism in one group makes racism more attractive in the other. The two groups' identity dynamics are jointly determined: each group's long-run share of racialists rises with the other group's racism. As shown in the sequel, the model indicates that locally stable steady states equilibria prevail with both individualists and racialists present in each community. An interesting insight from the framework is to highlight that the effect of social integration may be ambiguous: more intergroup contact can reduce racism when initial antagonism is low, but it can amplify racism when racist attitudes are already widespread, because more contact then exposes individuals more often to hostile out-group behavior.

More specifically, as in the socio-economic environment studied by [Darity Jr et al. \[2006\]](#), consider a large stationary population of agents is divided into two groups, indexed by $i = w, b$ and distinct by a racial characteristic, e.g., colour of skin, shape of eyes, etc. This characteristic is a phenotype that cannot be changed by deliberate choice, and defines their racial group. The size of community i is λ^i . Individuals however differentiate themselves also along their racial identity, which is adopted by choice: individualist I and racist R . Individualists live a race-free life, even though their social group is observable. Racialists, on the other hand, choose to identify strongly with their racial group. While the population shares of the two groups are fixed, individuals choose to adopt either an individualist or a racist identity. The within-group fraction - say of racialists - is therefore endogenous and evolves dynamically.

In each period, individuals are randomly matched in pairs and interact in productive socio-economic activities. Matching across racial groups is segmented so that the probabilities for an individual of group i to match with an within-group or an across-group individual are respectively:

$$Q^{ii} = \lambda^i + (1 - \lambda^i)(1 - s) \quad Q^{ij} = (1 - \lambda^i)s;$$

where s is a parameter capturing the degree of social segmentation between the two groups: when $s = 0$, the two groups are completely segmented, while conversely when $s = 1$, the groups meet according to random

matching.²² The productivity of the interactions depends on the mutual compatibility of the individuals identities. Racialists are altruistic toward members of their own social group, but antagonistic toward members of the other group. Individualists, on the other hand, are neither altruistic nor antagonistic toward any agent they interact with, socially.

Within a match, all individuals choose a discrete action : $x = 1$ - embracing openly a racist identity that differentiates between the two groups - or $x = 0$ and remaining race-blind. The material cost of undertaking the racist action is $C > 0$. More specifically, consider an individual h in social group i . When matching with a racist agent h' (that is, an agent choosing identity action $x = 1$), he/she obtains a positive payoff $B^i > 0$ if h' belongs to social group i , while the payoff is negative - equal to $-D^i$ - if h' is an out-group racist. The negative payoff is however reduced to $-D^i(1 - \delta)$ if agent h also chooses to adopt a racist identity $x = 1$. Finally, a racist identity also conveys a psychological payoff: an intrinsic benefit H^i as well as an added benefit R^i when matching with another in-group racist.

Formally, the productivity of a match (h, h') for an individualist h in social group i is:

$$U^{i,I}(x^h, x^{h'}) = \begin{cases} B^i x^{h'} - C x^h & \text{for } h' \text{ in } i \\ -D^i x^{h'}(1 - \delta x^h) - C x^h & \text{for } h' \text{ in } j \neq i \end{cases}.$$

For a racist h in social group i instead the payoff of a match (h, h') is:

$$U^{i,R}(x^h, x^{h'}) = \begin{cases} B^i x^{h'} - C x^h + (R^i x^{h'} + H^i) x^h & \text{for } h' \text{ in } i \\ -D^i x^{h'}(1 - \delta x^h) - C x^h + H^i x^h & \text{for } h' \text{ in } j \neq i \end{cases}.$$

Following [Darity Jr et al. \[2006\]](#), we assume $\delta D^i < C < H^i$; thereby guaranteeing that an individualist agent never takes a racist action while a racist always takes the racist action. It is then straightforward to obtain the equilibrium actions and utilities in each possible match.²³

Let q^i denote the fraction of racist agents in social group i . Assume that both the racist and individualist trait can be transmitted within each group from parents to children, according to the cultural transmission mechanism in [Bisin and Verdier \[2001b\]](#), in which parental socialization interacts with identity formation.²⁴ The parental preferences for socialization which motivate parents to transmit their own attitudes in their respective racial group take into account the probabilities of in-group and out-group matchings and hence are frequency dependent. More precisely, the parental paternalistic motivation $\Delta V^{i,I}(q^j)$ to transmit the individualist trait in a group depends negatively on the frequency of the racist trait q^j in the other group; and this is more so the more frequent the across-group matches Q^{ij} and the larger are the negative consequences of racist actions on the targeted community - that is, the larger is D^i .

On the other hand, the paternalistic motivations $\Delta V^{i,R}(q^i, q^j)$ to transmit the racist trait in a social group depends positively on the frequency of the racist in both social groups. Indeed, first of all, the larger the frequency of in-group racists q^i , the more likely a racist gets, through in-group contacts, the psychological gain R^i from sharing and expressing jointly the same racial identity. This effect is stronger the larger the frequency of in-group contact Q^{ii} . Further, the larger the frequency of out-group racists

²²[Darity Jr et al. \[2006\]](#) allows for s to depends on the frequency of racists in both the racial groups.

²³Appendix 1 provides all the formal derivations of the model.

²⁴For simplicity, we assume that there are no cross-group cultural influences; see [Bisin et al. \[2011\]](#) for a framework allowing such influences.

q^j , the more useful is the racist identity in matches with out-group individuals, as it has defensive value δ against the racist action of the out-group. This effect is stronger, the larger the frequency of intergroup contacts Q^{ij} .

Given the shape of the paternalistic motivations to transmit cultural traits, deploying the mechanics of the cultural transmission process in Bisin and Verdier [2001b] - under standard regularity conditions, including quadratic parental socialization costs $\phi(d) = \varphi \frac{d^2}{2}$ - we obtain the cultural dynamics of racist identities inside each racial groups:

$$\dot{q}^i = q^i(1 - q^i)\varphi [\Delta V^{i,R}(q^i, q^j)(1 - q^i) - \Delta V^{i,I}(q^j)q^i] \quad \text{for } i, j \in \{w, b\} \quad ; \quad i \neq j. \quad (7)$$

To better illustrate these cultural dynamics, we now construct the phase diagram of this system of differential equations; see Figure 1. An interior cultural steady state within community $i \in \{w, b\}$, given a frequency q^j of outgroup racials $j \neq i$, is characterized by the following condition:

$$\frac{q^i}{1 - q^i} = \frac{\Delta V^{i,R}(q^i, q^j)}{\Delta V^{i,I}(q^j)}. \quad (8)$$

Solving (8) provides steady state cultural manifold of racials in group i as a function of the frequency of outgroup racials q^j :

$$q^i = \tilde{q}^i(q^j) \in (0, 1) \quad \text{for } i, j \in \{b, w\} \quad \text{and } i \neq j$$

Importantly, $\tilde{q}^i(q^j)$ is increasing in q^j , as an increase of outgroup racials induces higher incentives to transmit the racist trait, and at the same time lower incentives to transmit the individualist trait within a given group. Hence, consistent with Darity Jr et al. [2006], racialism across communities is self-reinforcing: the diffusion of the racist trait inside a given group is favored by a higher frequency of racist out-groups.

The phase diagram for cultural dynamics of racialism in the space $(q^w, q^b) \in [0, 1]^2$ - in Figure 1 - indicates a unique locally stable interior cultural steady state $(q_*^w, q_*^b) \in (0, 1)^2$ at the intersection of the two cultural manifolds $q^w = \tilde{q}^w(q^b)$ and $q^b = \tilde{q}^b(q^w)$.²⁵

The comparative statics indicate that racialism is culturally more likely to be sustained when the emotional gains R^i and H^i of racialism increase, when the economic cost D^i associated to being subject to a racist action by an outgroup increases, and when the defensive value δ of racialism against across-group racialism increases. On the other hand, the impact of social integration s on the diffusion of racialism is less straightforward; see Figures 2a-2b. Indeed, it can be shown that an increase in social integration between the two communities tends to increase the steady state extent of racialism $q^i = \tilde{q}^i(q^j)$ within a group if and only the fraction of outgroup racials q^j is larger than a certain threshold $\bar{q}^j$.

Graphically, increased social integration makes the cultural steady state manifold $\tilde{q}^w(q^b)$ of group w to rotate clockwise around point W , and the cultural steady state manifold $\tilde{q}^b(q^w)$ of group b to rotate counter-clockwise around point B . The steady state equilibrium moves then from point E to point E' in Figures 2a-2b. This leads to increased racialism when racialism is already pervasive in both groups (Figure 2a), but instead it promotes reduced racialism when racialism is already relatively weak in

²⁵Compared to Darity Jr et al. [2006], the racist complementarity across groups does not lead to extremal steady states with a full individualist or full racist society. This is due to the cultural substitutability property associated to the cultural transmission process. This feature highlights the importance of the specificity of the micro-founded mechanism of cultural transmission for the pervasiveness of racist identities in a society.

Figure 1 : cultural dynamics of racialism

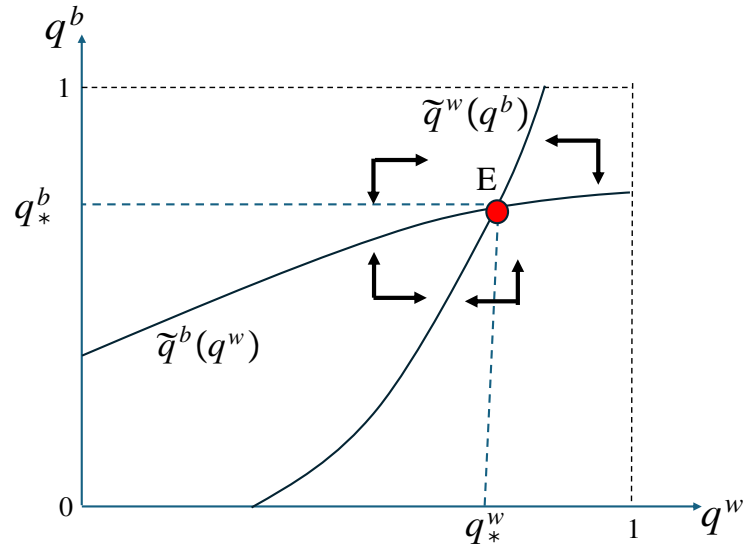


Figure 2a) : higher intergroup contact
high racialism

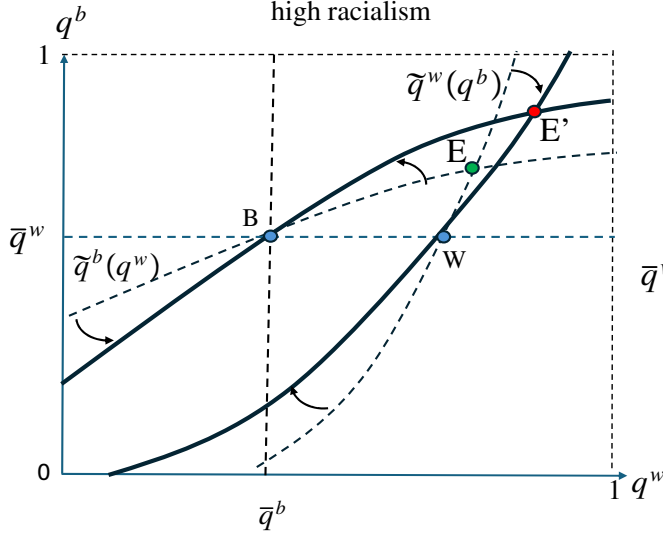
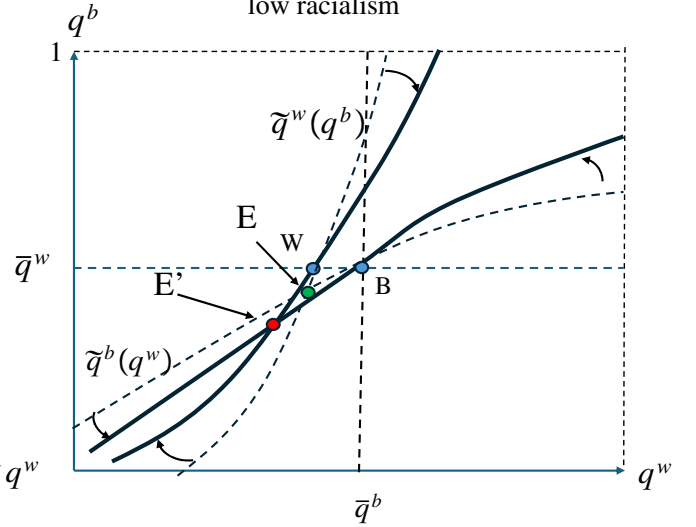


Figure 2b) : higher intergroup contact
low racialism



both communities (Figure 2b). In other words, social integration across communities reduces or exacer-

bate racist identities, depending on the initial degree of racial antagonism that exists between the two communities.²⁶

4.3 Cultural Polarization

Polarization across cultural traits - especially with reference to political ideologies and beliefs - is an important phenomenon, extensively studied and documented e.g., in political science and sociology. Within this context, [Michaeli and Wu \[2022\]](#) study whether societies can become more polarized over time even when peer effects in identity formation favor conformity. Specifically, [Michaeli and Wu \[2022\]](#) extend [Bisin and Verdier \[2001b\]](#) to allow for continuous cultural traits. Each parent is associated to a trait $i \in [0, 1]$ and each trait is associated to a archetypical action x^{i*} . Parents chooses which trait to socialize their children to, anticipating that they will choose an action that is socially relevant within peer interactions. A child with trait i chooses action x^i to minimize two forces: the *cognitive dissonance* cost $(x^i - x^{i*})^2$ of deviating from his/her own trait, and the *peer pressure* cost $\int (x^i - x^j)^2 \mu(dj)$, where μ is the distribution of cultural types in the population. Peer pressure therefore biases childrens' choices towards the population mean and each child's i optimal action is a simple average between the archetypical action associated to his/her trait x^{i*} and the mean action in the population $\int x^j \mu(dj)$. Parental socialization preferences are characterized by imperfect empathy: they face a psychological costs from their children choosing an action which is different from their (the parents') archetypal action. Parental socialization preferences are however also modulated by their internalization of the social and psychological costs the children may bear from peer pressure. This component of parental preferences is weighted by a parameter $\alpha > 0$. A higher value of α means parents care more about reducing their children's disutility from peer-pressure rather than socializing them toward their own preferred behavior, leading therefore the cultural dynamics towards homogenization overtime. Conversely for low values of α , parental socialization preferences may induce their socialization effort towards traits farther away from the parents', and cultural polarization in the long run.

[Michaeli and Wu \[2022\]](#) then extend their model to allow parents to choose neighborhoods as an alternative response to peer pressure effects. In such a case, children face peer pressure within neighborhoods only. In equilibrium, neighborhoods sort by trait: low types clustering in a given place, and high types clustering in another place. When α is large enough, parental internalization produces within-neighborhood convergence, but now the long-run outcome involves the formation of two homogeneous neighborhoods with different norms. One gets reduced within-group variance but persistent between-group differences. In other words, segregation blunts the integrative force of parental internalization, though it can still prevent the most extreme polarization forms relative to the no-internalization benchmark.

4.4 Identity Switching, Assimilation and Cultural Openness

Identity formation on the part of children might involve switching, that is, turning to a different cultural trait than the one they have been socialized to by parents. This phenomenon is studied by [Giavazzi et al. \[2019\]](#). In this paper, as in [Bisin et al. \[2011\]](#), parents first choose the family socialization rates and

²⁶This result aligns with the *contact hypothesis* in social psychology [[Allport, 1954](#)], which posits that interactions between members of different groups can improve intergroup relations, only when specific conditions are met, that is, meaningful contact challenges stereotypes, but successful outcomes depend heavily on specific facilitating conditions such as equal status, common goals, intergroup cooperation, and institutional support.

then children choose their own cultural identity. The specific identity formation mechanism has children meeting randomly with other individuals in the population. But prior to meeting, each child forms his/her identity trading off the costs of switching cultural identity with the associated gains. Let V denote the value of a match between individuals with the same cultural trait. Let θ^M (resp. θ^m) denote the fraction of the value of the match lost when a child keeps (resp. switches from) the parental identity and meets with an individual of the other (resp. parents') group. Switching cultural identity involves a direct cost for the child, $\rho \cdot d + z$, in part controlled by parental socialization effort d and in part stochastic, as determined by the realization of z , uniformly distributed in $[0, \bar{z}]$. A child switches identity if the expected gain is higher than the expected gain from keeping the parental identity:

$$(1 - q)V\theta^M - q\theta^m V - \rho d - z \geq 0.$$

While parents - with imperfect empathy - derive utility from transmitting their own cultural trait, they also care about the usefulness of that trait for their children when they meet in society. They choose their optimal level of socialization taking into account the child's own optimal choice problem, and knowing the distribution of the switching cost, but not the particular realization for their child. [Giavazzi et al. \[2019\]](#) specialize the model to one where the cultural choice is only on the part of the minority group i - and switching may be interpreted as assimilating. The socialization rate d chosen by parents is shown to be increasing in ΔV^i , decreasing in the cost of the effort. Most interestingly, d is decreasing in θ^M and increasing in θ^m . Altruistic parents in fact internalize the cost ρd imposes on their children and reduce their socialization effort the larger is θ^M and decrease it the smaller is θ^m .

Two types of stationary states may emerge in this model: one where all dynamical paths converge to assimilation of the minority; and another where the minority group does not assimilate. The occurrence of complete assimilation depends upon a set of parameters such as the child's transaction gains and switching costs from assimilating, the degree of intolerance of parents, the technology of parental socialization, and finally on how much parents account for their child's utility payoff, that is, their altruism. All these parameters are likely to vary across different types of cultural traits. Furthermore, and very importantly, [Giavazzi et al. \[2019\]](#) obtain results about the speed of convergence as a function of the parameters of the environment, identifying the characteristics of traits which are likely to spread quickly across generations.²⁷

Along a similar vein, [Itoh et al. \[2024\]](#) develops a dynamic theory of minority assimilation in which the effort of assimilation increases income, but also affects "perceived distance" from the majority norm, and influences relative group-status. The model features an explicit intergenerational linkage by assuming that assimilation by one generation lowers the effort cost for the next generation. As a consequence, assimilation can persist and accelerate over time. Similar to [Giavazzi et al. \[2019\]](#), the model may generate multiple stable steady states, specifically when the perceived-distance motive is strong enough relative to the material/status return to effort. This is so because in the model's specification, perceived distance is assumed to be lowest at "no assimilation" and "full assimilation" and highest at intermediate effort, making incentives non-concave and allowing both low-assimilation and high-assimilation long-run outcomes to coexist for intermediate parameter values. The paper then endogenizes the majority group as an active player choosing a costly level of "tolerance" toward the minority. Higher tolerance reduces the majority's disutility from interacting with culturally distant minorities and, in the aggregate, lowers minority assimilation costs. The paper shows that a hump-shaped relationship between tolerance and assimilation prevails.

²⁷These theoretical predictions are confirmed by an empirical analysis mapping the speed of evolution of a wide range of cultural attitudes for different generations of immigrants to the United States.

At low levels of assimilation, increased minority effort tends to increase majority tolerance. However, at high levels of assimilation, further effort can trigger a "competition or threat" response, causing majority tolerance to decline. From a normative point of view, the analysis highlights both intragenerational and intergenerational externalities across the two groups. As a result in steady state, majority tolerance is typically lower than the social optimum, while minority effort may be too low or too high depending on how they undervalue future generations, or how they ignore the social friction caused to their own peers. Finally, from a policy perspective, the paper suggests that to reach a given assimilation level, an indirect policy that increases majority tolerance should be preferred to a direct policy that pushes minority assimilation effort. Indeed tolerance-focused policies generate broader intergenerational benefits and reduce the "identity friction" that direct assimilation policy often exacerbate.

Thill and Zana [2026] also studies an environment in which children can switch utility - after parental socialization - while they engage in social interactions with their peers. In the formulation in the paper, parents culturally transmit a trait of *cultural openness*, an attitude relatively favorable to their adoption of a new cultural trait when beneficial. Identity formation entails a choice about how much to interact with other groups and, conditional on interaction, whether to switch traits. More specifically, in Thill and Zana [2026] the parental cultural trait i is directly transmitted to the children; and parental socialization effort affect their openness - that is, probability that they will switch trait. The decision whether to keep or switch - the identity formation of the children - is taken as they engage in various forms of social interactions. Interactions among individuals within the same cultural group reinforce existing traits, while interactions with different cultural groups open the option of switching, at a cost representing, e.g., social stigma. Each group has an associated level of tightness, defined as the strength of stigma associated with leaving the group. Cultural openness - the propensity to adopt new cultural traits - effectively reduces these costs. This, in turn, shapes how likely their child is to change their trait and hence the likelihood for cultural continuity across generations. The parental motivation to socialization trades-off - similarly to Michaeli and Wu [2022] - imperfect empathy with the costs faced by children when switching and adopting alternative cultural traits. Interestingly, cultural openness has a negative effect on the dynamics of the population share of a cultural trait, while tightness of the stigma has counterbalancing effects: it reduces exits from the cultural group but it also reduces entry, by making the group less attractive to outsiders. Depending on parameters, the cultural dynamics can lead to the survival of a single cultural group, to the coexistence of both groups, with limited interactions and switching between them, or even the segregation of the groups, with no interactions nor switching.

5 Social Identities at the Extensive Margin

In the previous section we surveyed models in which identity formation is embedded within inter-generational cultural transmission process. Specific identity formation problems are studied in detail, like e.g., oppositional and racist identities. In these models identity formation choices by children are restricted to the set of existing cultural traits in the population and hence generally operate at the intensive margin - e.g., adopting oppositional attitudes within minoritarian cultures or racist attitudes within groups defined by given racial characteristics. Even when children can switch the cultural trait they share with their parents - they can just switch to the other trait in the population. In this section we instead study identity formation in environments in which children choose identities constructed as a rich set of possible

combinations of a vector of characteristics.

A general approach to identity formation along this lines has been introduced by [Shayo \[2009\]](#). Individuals with fixed individual characteristics choose a social identity trading-off how they fit within the social group and the social esteem or social status that belonging to that group confers. Specifically, the population is composed of two cultural groups i , whereby i represents a vector of fixed attributes. Let q^i denote the fraction of agents with attributes i .

An individual with attributes i takes a social action x^i generating an individual payoff $\pi^i(x^i, X)$, where X is an aggregator of individual actions: e.g., $X = \sum_i q^i x^i$. Individuals in group i choose to identify to a social group $g \in G = \{1, \dots, G\}$, where g represents the archetype vector of attributes of the group. Identifying to g , for an individual i induces a psychological cost related to the social distance between the groups' characteristics $\delta(i, g)$. At the same time, identification confers some social esteem related to the social status of the group $s^g = s(\pi^g, x^g, q^g)$, where π^g, x^g, q^g denote, respectively, the average profits, actions, relative size of the group g . The utility function of an individual with attributes i who identifies to g is

$$U^i(x^i, X; g) = \pi^i(x^i, X) - \delta(i, g) + \alpha s(\pi^g, x^g, q^g);$$

where $\alpha > 0$ reflects the relative importance of social status with respect to cognitive distance on the individual's preferences.

Individuals with attributes i choose first the group(s) they identify to g ; and in a second stage they choose their optimal action x^i . In both stages individual take for given the map $G(i)$ and the second state choices of all individuals, $x^{G(i)}$, for all characteristics i , not just the individual's own.²⁸ By backward induction, consider the choice of action of an individuals' i who identifies to g and takes X as given:

$$\begin{aligned} & \max_{x^i} U^i(x^i, X; g) \\ \text{s.t. } & \pi^g = \sum_{g=G(j)} \pi^j, \quad x^g = \sum_{g=G(j)} x^j, \quad q^g = \sum_{g=G(j)} q^j. \end{aligned}$$

The solution $x^i(X, G)$ at equilibrium must satisfy the fixed point condition

$$X = \sum_i q^i x^i(X, G).$$

Consequently, with enough regularity conditions, individual equilibrium choices $x^i(G)$ are determined. The identification choice of individuals with characteristics i is then written as:

$$\max_g U^i(x^i(G), X; g),$$

and it induces a map $g = F(i, G)$. At equilibrium

$$G(i) = F(i, G), \text{ for all } i.$$

The resulting equilibrium pattern of identity choices depends on the collection of attributes i and their population distribution q^i .

²⁸Formally, we are formulating an atomistic game, where the population is composed of a mass of individuals for any characteristic i .

While the identity formation approach of [Shayo \[2009\]](#) we just introduced is purely intra-generational, identity depends on the equilibrium population shares q^i and hence can naturally be embedded into the dynamics of the inter-generational cultural transmission model in [Bisin and Verdier \[2001b\]](#). These dynamics now will depend on $G(i)$ which in turn depends on q^i , inducing endogenous interactions in parental preferences for socialization. We refer the reader to the Appendix for this analysis and turn instead to an interesting application in the next section.

5.1 Broad Versus Narrow Identities, and Cultural Transmission

In a given society, individuals generally hold multidimensional identities, e.g., ethnicity, class, religion, tastes, geographic, etc. Furthermore, they might identify with multiple social groups rather than a single one [Dasgupta and Goyal \[2019\]](#), [Sen \[2013\]](#). The sociological literature has investigated panethnic identity formation — that is, individuals choosing whether to identify narrowly (Chinese, Filipino, Korean; Mexican, Cuban, Salvadoran) or broadly (Asian American, Latino/Hispanic) with the broad group being an aggregation created partly by government racial categorization and partly by political coalition-building; see e.g., [Lê Espiritu \[2019\]](#) foundational work. Interestingly, [Martínez and Gonzalez \[2021\]](#) found that Latino panethnic identification functions substantially as a distinction vs. status tradeoff structure: people adopt the broad "Latino/Hispanic" label more strongly in response to external threat or discrimination. Also, studies of residential patterns - e.g., [Kim and White \[2010\]](#) - find real panethnic clustering (Latinos, Asians living closer to co-panethnics than to other groups) but also persistent ethno-national distinctiveness within those categories.

In this context, it is then of interest to characterize conditions under which individuals adopt narrow identities - e.g., organized around high-salience markers like religion, ethnicity, or race - or else conditions which favor the adoption of broader identities. To address this issue [Dasgupta and Goyal \[2019\]](#) study a *club-theory* model whereby identity emerges from the interaction of individual incentives, group-level interests, and spillovers between groups. Relatedly, [Carvalho \[2021\]](#) introduces an identity formation model in which narrow and broad identities are the result of cognitive responses to the salience of various characteristics that individuals may use to define their sense of oneself. In both cases, the discussed mechanisms are purely intra-generational, that is, taking as given the repertoire of cultural characteristics (or traits) that the society has inherited from the past. Taking a longer perspective, one may however expect that the cultural repertoire individuals have access to also evolves overtime, affecting therefore dynamically the process of identity formation within generations.

Conversely, as suggested in our general framework in Section 3, the intergenerational process of cultural transmission is likely to be affected by the specific degree of broadness of identity chosen in each generation. Typically, in standard economic models of cultural transmission, one generally assumes a one-to-one fixed correspondence between individual-level cultural traits and group level characteristics. Extensive margin social identity formation processes (ie. the endogenous mixing of cultural traits that define a specific group identity structure) breaks this fixed relationship. Indeed a person may identify narrowly with her own cultural group, such as ethnicity, religion, class, or region, or broadly with a more inclusive group that contains several traits. This feature in turn is likely to affect the transmission mechanisms of these cultural traits across generations.

In this section, we map the issue of broad versus narrow identities into the identity formation model delineated in the previous section, following [Shayo \[2009\]](#). To this end, we let the set of social groups G to

include groups composed of collections of cultural traits, to capture broad identities. We then embed this identity formation environment into the cultural transmission model of [Bisin and Verdier \[2001b\]](#). The setup has two cultural groups (a) and (b). Individuals choose effort, which generates output, and then choose whether to identify narrowly with their own group or broadly with the inclusive group (ab). Identity choice trades off two forces: a psychological distance cost from identifying with a group that is less close to one's inherited trait, and a social-esteem benefit from belonging to a higher-status or more productive group. Hence, broad identity becomes more attractive when the broad group is sufficiently close to one's own trait and when social esteem is sufficiently salient relative to cultural distance. Different identity patterns emerge in equilibrium: both groups may remain narrow, both may adopt a broad identity, or one group may be narrow while the other becomes broad.

A central result is that identity breadth feeds back into cultural transmission. Parents with a narrow identity have stronger incentives to transmit their own cultural trait, because their self-understanding is closely tied to that trait. Parents with a broad identity transmit their own trait less intensely, because their identity is less exclusively attached to it; moreover, the incentive to transmit a broad-identity trait depends positively on how common that trait already is. This generates different cultural steady states across identity regimes. With cultural substitutability in socialization, one gets a unique stable interior steady state within each regime, but the location of that steady state differs depending on whether identities are narrow, broad, or mixed.

Interestingly, as the salience of social esteem rises, the analysis shows that society moves from a full narrow-identity regime to mixed regimes and eventually to a full broad-identity regime. Also, the effect on frequency of the minority trait (a) is non-monotonic. At intermediate values of the salience of social esteem, the majority trait (b) adopts the broad identity first, weakening its own transmission advantage and raising the long-run share of the minority trait. But when the salience of social status is high enough for both groups to adopt broad identities, the broad-identity mechanism tends to favor the already larger group, reducing the minority share again.

Formally, consider a society whose population is composed of two cultural groups characterized by a vector of fixed attributes. The groups are indexed by $i \in \{a, b\}$ and q^i denotes the fraction of group i in the population. Individuals i choose effort x^i which generates output $y^i = Rx^i$, with $R > 0$ denoting the return to effort. Individuals are differentiated by their disutility of effort. Specifically an individual i has a disutility of effort $c^i(x) = \theta^i x^2/2$. Let the payoff arising from x^i be $\pi^i(x^i) = Rx^i - c^i(x^i)$. After acquiring their trait through parental socialization, children choose to identify to a social group $g \in G$. Specifically, the set G contains three groups $G = \{\{a\}; \{b\}; \{ab\}\}$: the two cultural groups indexed by i and a broad group $\{ab\}$. In other words, a child i can choose to identify narrowly with his/her own group i or else he/she could identify with a broad social identity - group $\{ab\}$ - inclusive of all individuals independently from the cultural trait inherited through direct parental socialization.²⁹

As in our analysis of [Shayo \[2009\]](#) in the previous section, we assume that the choice of social identity for an i individual trades-off the psychological cost related to the social distance between i and the the group g whose identity is adopted - $\delta(i, g)$ - and the social esteem gains associated to group g - s^g . Specifically we formulate social esteem s^g as the difference $Y^g - Y^{-g}$ between the income of group g and its reference group $-g$.³⁰ More precisely, the utility to adopt identity $g \in G$ for individuals with characteristic i is given

²⁹The agent i could in principle choose to identify narrowly with the other social group, but will never choose to do so under straightforward regularity assumptions.

³⁰Hence $Y^i = q^i y^i$; $Y^{-i} = Y^j = q^j y^j$ for $i, j = a, b$, $i \neq j$, $Y^{ab} = q^i y^i + (1 - q^i) y^j$ and, for convenience, we normalize

by:

$$U^i(x^i, g) = \pi^i(x^i) - \beta(\delta(i, g)) + \gamma s^g;$$

where

$$\delta(i, g) = \begin{cases} = 0 & \text{if } g = i \\ = 1 & \text{if } g = j, \\ = q^j, & \text{if } g = ab \end{cases}, \quad s^g = \begin{cases} q^i y^i - (1 - q^i) y^j, & \text{if } g = i \\ q^i y^i + (1 - q^i) y^j, & \text{if } g = ab \end{cases};$$

and where β and γ reflect, respectively, the salience of the cultural characteristic i and the salience of social esteem.

Identity formation choice for an individual i has the same timing as in the general process delineated in the previous section: he/she first chooses social identity - either narrow identity $g = i$ or broad $g = ab$; and then effort x^i .

It is straightforward to solve for the identity formation process. Individual i chooses effort $x^i = \frac{R}{\theta^i}$ and hence he/she chooses the narrow identity i if and only if:

$$q^i < \hat{q}^i = 1 - \frac{2\gamma}{\beta} y^j. \quad (9)$$

The pattern of identity choice in the society depends on the distribution of cultural traits q^i , as well as on the salience parameters of identity choices. A cultural type i chooses a broad social identity b if and only if his/her trait i is sufficiently represented in the society. In such a case, this individual feels sufficiently close to the cultural archetype of the society at large to choose to take the broad identity rather than the narrow identity exclusively focused on his social group. Also the larger is the relative salience of social status compared to social distance γ/β , the more likely the broad social identity will be chosen by any cultural type. Specifically, it can be shown that for γ/β sufficiently small and for q^i in some interval - that is, for a population sufficiently balanced - both groups adopt a narrow social identity associated to their own trait. Conversely, when γ/β is large and the population again balanced, both groups identify with the broad social identity. The pattern of identity choice for the two social groups is depicted in Figure 3. In the first case, $\hat{q}^a > 1 - \hat{q}^b$, there is no overlap of broad identity choices by the two cultural groups. In the second case, $\hat{q}^a < 1 - \hat{q}^b$, there is an intermediate region in which the two cultural groups adopt jointly a broad identity structure.

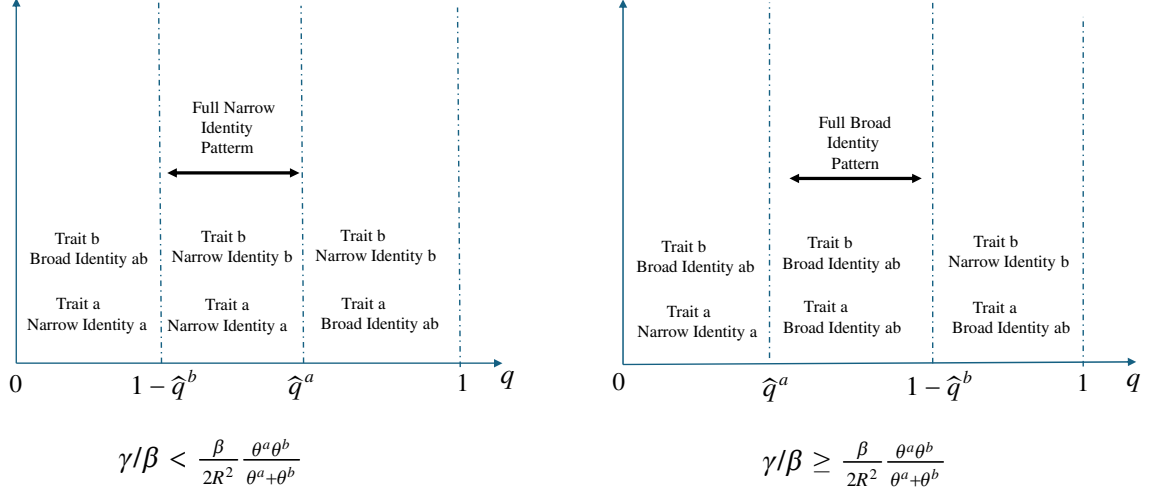
We can now study the cultural transmission process in this socio-economic environment. Interestingly, while the distribution of the population by cultural trait q^i affects social identity choices, as it is generally the case, the identity choice of parents affects parental socialization choices and hence the pattern of cultural evolution. This is the case inasmuch as the parental preferences for socialization depend not only on the anticipated effort choices of the children - as determined by imperfect empathy - but also on the distance between the attributes of the parent and the child. Letting ΔV_i^i denote the socialization preferences of a parent i with narrow identity i and ΔV_b^i the socialization preferences of a parent i with broad identity ab :³¹

$$\Delta V_i^i = \frac{R^2}{2\theta^j} \frac{[\theta^j - \theta^i]^2}{\theta^i \theta_j} + \beta; \quad \Delta V_{ab}^i = \frac{R^2}{2\theta^j} \frac{[\theta^j - \theta^i]^2}{\theta^i \theta_j} + \beta [2q^i - 1] \quad (10)$$

$Y^{-ab} = 0$.

³¹See the Appendix for derivations; we assume for simplicity that parents do not account for the identity choice of their children in their parental preferences for socialization.

Figure 3: Patterns of identity choices



We conclude then that $\Delta V_{ab}^i \leq \Delta V_i^i$. Namely individuals with a broad social identity have lower incentives to transmit their cultural trait than individuals with a narrow identity. Moreover, ΔV_{ab}^i is increasing in q^i ; that is, for individuals with broad identities, the preferences to transmit their own trait are increasing in the frequency of that trait.

Furthermore, we assume without loss of generality that $i = a, b$ and $\theta^a < \theta^b$. We refer to trait a as the minoritarian trait.³² Letting $q_t = q_t^a = 1 - q_t^b$, we can solve for the optimal parental socialization rates $d^a(q_t)$, and $d^b(q_t)$. Importantly, as depicted in Figure 4 the shape of the socialization rate $d^i(q)$ of a parent with trait i depends on the identity regime $g(i) \in \{i, ab\}$ of such parent, which in turn is affected by how q_t compares to the thresholds $\hat{q}^a$ and $1 - \hat{q}^b$; see equation 9.

The cultural dynamics are represented by the following standard equation:

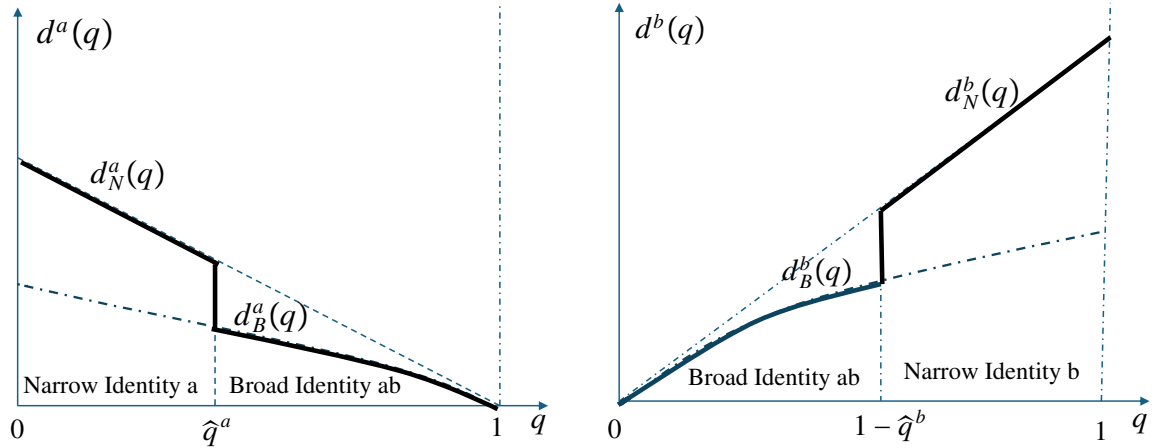
$$q_{t+1} - q_t = q_t(1 - q_t) \left(d^a(q_t) - d^b(q_t) \right). \quad (11)$$

Under the cultural substitutability property of the socialization rates,³³ there is a unique interior and stable cultural steady state. Social identity within generations (narrow or broad) can then be in one of four possible regimes, at any given point of time: i) a full narrow regime, in which both cultural types a and b adopt a narrow identity anchored on their own trait; ii) a full broad regime, in which both cultural

³²In this case - and in an environment with no identity formation choice, in which e.g., $\beta = \gamma = 0$ - the stationary state would be $q = \frac{\theta^a}{\theta^a + \theta^b} < \frac{1}{2}$.

³³In the appendix, we provide sufficient conditions ensuring that both $d^a(q_t)$ and $d^b(q)$ belong to the interval $[0, 1]$, and satisfy the "cultural substitutability" property, namely that $d^a(q)$ (respectively $d^b(q)$) is decreasing (respectively increasing) in $q \in [0, 1]$.

Figure 4: Socialization rates of trait a and b and identity choices



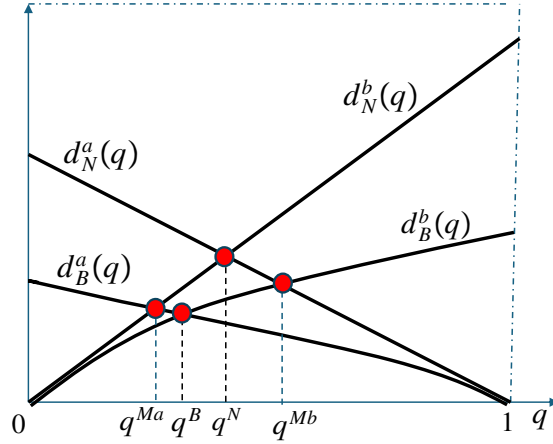
types adopt a broad identity inclusive of both traits; iii) a mixed regime with narrow identity for trait a and broad identity for trait b ; iv) a mixed regime with narrow identity for trait b and broad identity for trait a . Stationary states can belong to any of the four regimes, let q^a in each of them, respectively, be denoted q^N ; q^B ; q^{Mb} ; q^{Ma} .

The cultural steady states under the different identity regimes are shown in Figure 5. The figure shows that $q^{Ma} < q^B < q^N < q^{Mb}$.³⁴ This ranking reflects two features. First, as shown above in Equation (10), an individual is less likely to socialize his/her child to his own trait when he embraces a broad identity. As a consequence, that trait is at a cultural selection disadvantage when the opposite trait is transmitted by individuals adopting a narrow identity structure. Secondly, adopting a broad identity creates for the paternal socialization preferences a component of strategic complementarity. This second effect is relatively stronger for the majoritarian cultural trait b than for the minoritarian trait a . As a consequence when both adopt a broad identity, the minoritarian trait is at a relative disadvantage in terms of cultural transmission, and the full broad regime induces a lower cultural steady state q^B of the minoritarian trait a compared to cultural steady state of the full narrow regime q^N . Compared that full narrow regime, the majoritarian trait is then at a disadvantage if he is the only to adopt a broad identity (ie. $q^N < q^{Mb}$). Conversely, compared to the full broad regime, the minoritarian is also at a disadvantage if he is the only one to adopt a broad identity (ie. $q^{Ma} < q^B$).

The cultural dynamics depend then on the identity regime in which society is, which in turn is affected

³⁴See the appendix for a derivation.

Figure 5: Cultural steady states in different identity regimes

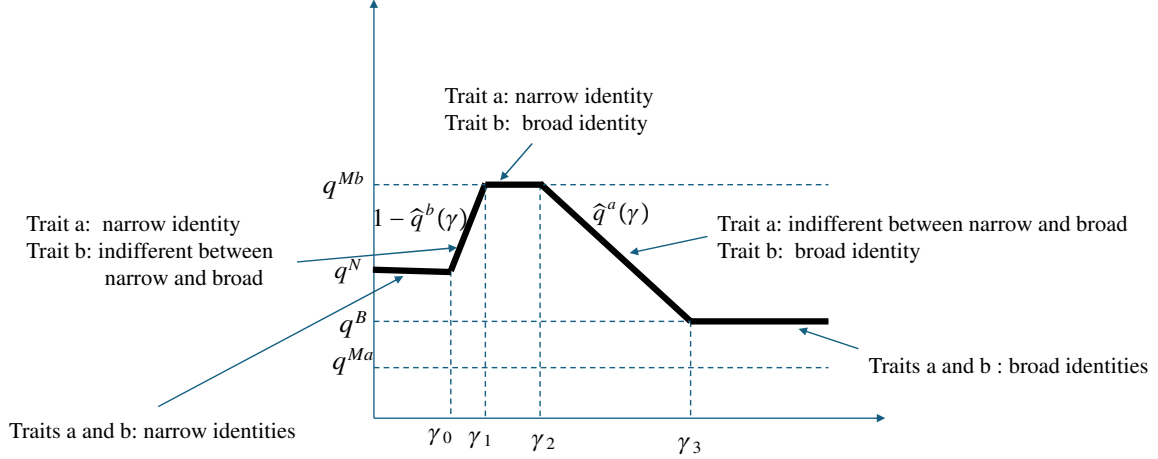


by the distribution of the population by cultural trait. Typically, different long term situations arise depending on the identity regime at which the socialization rates $d^a(q)$ and $d^b(q)$ intersect. Inspection of Figure 6 shows that this relates to the relative position of the identity thresholds $\hat{q}^a$ and $1 - \hat{q}^b$ as well as to the different steady states q^{Ma} , q^B , q^N , and q^{Mb} .

A complete characterization of the cultural dynamics associated to identity choices can be obtained as a function of the salience γ , the social prestige component of identity choices, as illustrated in Figure 6.³⁵ Four long term cultural and identity regimes are obtained: for γ low enough (ie. $\gamma < \gamma_0$), a full narrow identity regime prevails with a steady state cultural population characterized by q^N . On the opposite when γ is large enough (ie. $\gamma > \gamma_3$), a full broad identity regime obtains with a long term cultural mix associated to q^B . Finally, for intermediate values of γ (ie. $\gamma \in [\gamma_0, \gamma_3]$), society ends up in a mixed regime

³⁵Indeed from equation (9), one can note that $\hat{q}^a$ is increasing in γ , while $1 - \hat{q}^b$ is decreasing in γ . At the same time, the identity regimes cultural steady states q^{Ma} , q^B , q^N and q^{Mb} are independent from γ . From this one can fully parametrize the long term cultural dynamics as a function of γ .

Figure 6: Cultural steady states and long run identity choices



where the minoritarian cultural type is likely to choose a narrow identity and the majoritarian type tends to adopt the broad identity pattern.

Interestingly, the steady state cultural trait distribution of the population changes non-monotonically with γ . Indeed, the larger γ , the more likely that an individual will adopt a broad identity. But individuals with the majoritarian trait *b* will adopt a broad identity at a lower threshold of γ . As γ increases, therefore, cultural trait *b* is first at a disadvantage in terms of cultural transmission, increasing therefore the long run frequency of trait *a* in the population. For values of γ large enough individuals *a* also adopt the broad identity. Given the strategic cultural complementarity component associated to identity choices, this re-establishes a cultural transmission advantage towards the majoritarian trait *b*.

6 Conclusions and Future Research

In this chapter we aimed at integrating the theory of identity formation within population dynamics models of the evolution of culture. Conceptually, this can be viewed as integrating the literature on economic cultural transmission in Bisin and Verdier [2025, 2001b, 2010] with the literature on social identity in Akerlof and Kranton [2000, 2005, 2010]. While the canonical modeling of cultural evolution treat the children's choices about their own identity mechanically, we have surveyed existing research that centers on the interaction between parental socialization choices and children identity formation. This research has demonstrated its explanatory power by empirically addressing the relative role of cultural conformity and cultural distinction in the formation of oppositional identities - as we discuss in in Section 4.1. Oppositional identity formation indeed represents a relevant social phenomenon in the political economy of race and of immigrant cultural integration.

Futhermore, in this chapter, we have extended fundamental existing work on identity formation - like [Darity Jr et al. \[2006\]](#) on racialism, [Shayo \[2009\]](#) on general social identity, and [Dasgupta and Goyal \[2019\]](#) on multidimensional identities - to formally integrate it into dynamic models of cultural evolution.

Clearly, several extensions could be pursued along these lines. Typically in the models we surveyed, intra-generational socio-economic interactions between identity groups are modeled in a static manner. Yet one of the central insights of the stratification economics literature is that patterns of racial discrimination across groups have dynamic consequences for the accumulation of human and physical capital by individuals ([Chelwa et al. \[2022\]](#), [Darity \[2022\]](#), [Davis \[2015, 2019\]](#), [Darity et al. \[2022\]](#), [Boerma and Karabarbounis \[2023\]](#)).³⁶ Unequal accumulation paths, in turn, feed back into subsequent discrimination and into the persistence of racialized cultural patterns across groups and generations. Incorporating such dynamic considerations into existing models of identity formation and cultural evolution could therefore prove highly fruitful for understanding the interplay between social stratification and the dynamics of identity formation.

Similarly, when identity and cultural traits are affecting localized socio-economic interactions and externalities, spatial economic differentiation may arise as a natural outcome if individuals can choose where to live or sort themselves ([Bénabou \[1993\]](#) [Hoff and Sen \[2005\]](#), [Picard and Zenou \[2018\]](#), [Helsley and Zenou \[2014\]](#)). This, in turn, provides a rationale for the emergence of local identity patterns that are self-sustaining and reinforce localized processes of cultural transmission. Introducing an explicit spatial dimension into the frameworks surveyed here could thus help clarify the role of space in the emergence and persistence of identity and cultures across groups.

As well, a central feature of economic models of intergenerational cultural transmission is the assumption that parents devote resources to socializing their children into specific cultural traits. In these frameworks, transmission is typically represented through a technology in which parental socialization efforts interact with mechanisms of exposure to cultural role models, whether vertical, oblique, or horizontal. In this regard, building on insights from developmental psychology ([Baumrind \[1971\]](#)), the recent economic literature on parenting styles ([Doepke and Zilibotti \[2017, 2019a,b\]](#)), has usefully emphasized the role of parenting practices—authoritarian, permissive, and authoritative—in shaping this technology of intergenerational transmission. Extending this perspective to incorporate the dimension of intragen-erational identity choices could deepen our understanding of when and how particular patterns of social identity become associated with specific parenting practices, how these associations interact, and what their consequences are for long-run cultural evolution in a given society.

Another promising line of inquiry concerns the relationship between identity formation, cultural change, and institutional structures. A recent literature in economics has emphasized the importance of interactions between culture and institutions for deepening our understanding of political economy, development outcomes, and the historical trajectories of societies ([Alesina and Giuliano \[2015\]](#); [Acemoglu and Robinson \[2024\]](#); [Bisin and Verdier \[2024\]](#)). Embedding the models developed in this paper within frameworks of institutional change may therefore be essential for highlighting the long-run connections between power, institutions and identity formation processes.

Besides the theoretical advancement, these extensions are important in our view because these models can be taken to data - structurally - to study several fundamental socio-economic phenomena. Once again, the issues of race and of the cultural and economic integration of immigrants constitute first order

³⁶See also [Brundage et al. \[2026\]](#) for a recent survey on the theoretical issues.

application of the theoretical analysis. First of all, these models have a role in better understanding the dynamics of race relationships, especially in the U.S. where, e.g., affirmative action policies seem to have had unanticipated effects regarding identity formation on either side of the racial divide. Furthermore, the recent surge in migration flows into Western countries is at the core of shattering reallocation of political power along the immigration policy divide both in Europe and in the U.S. The perceived externalities immigration imposes on natives and the perceived sluggishness of the cultural integration process of immigrants are the center motivation of these phenomena. While empirical analyses of cultural dynamics in this context exist [[Bisin and Tura, 2019](#)], by not centering on identity formation they may miss specific dimensions of second and third generation dynamics.

We believe that the explanatory power of models of identity formation embedded into intergenerational cultural transmission frameworks can be of great use in understanding population dynamics phenomena of socio-economic importance across the cultural dimension in future decades. We hope that such models will also help designing effective and efficient policies in this respect.

References

- Dominic Abrams and Michael A Hogg. Comments on the motivational status of self-esteem in social identity and intergroup discrimination. *European journal of social psychology*, 18(4):317–334, 1988.
- Scott F. Abramson and Moses Shayo. International integration and social identity. *American Economic Review*, 2022.
- Daron Acemoglu and James A Robinson. Culture, institutions and social equilibria: A framework. *Journal of Economic Literature*, forthcoming, 2024.
- G. A. Akerlof and R. E. Kranton. Economics and identity. *Quarterly Journal of Economics*, 115(3): 715–753, 2000.
- George A. Akerlof and Rachel E. Kranton. Identity and schooling: Some lessons for the economics of education. *Journal of Economic Literature*, 40(4):1167–1201, 2002.
- George A Akerlof and Rachel E Kranton. Identity and the economics of organizations. *Journal of Economic perspectives*, 19(1):9–32, 2005.
- George A Akerlof and Rachel E Kranton. Identity economics: How our identities shape our work, wages, and well-being. In *Identity Economics*. Princeton University Press, 2010.
- A. Alesina and P. Giuliano. Culture and institutions. *Journal of Economic Literature*, 53(4):898–944, 2015.
- Floyd H Allport. The structuring of events: outline of a general theory with applications to psychology. *Psychological Review*, 61(5):281, 1954.
- G. W. Allport. The psychological nature of personality. *The Personalist*, 34(4):347–357, 1953.
- Solomon E. Asch. Opinions and social pressure. In *Scientific American*, volume 193, pages 31–35. Scientific American, Inc., 1955.
- David Austen-Smith and Roland G Fryer Jr. An economic analysis of “acting white”. *The Quarterly Journal of Economics*, 120(2):551–583, 2005.
- Harinder Battu, McDonald Mwale, and Yves Zenou. Oppositional identities and the labor market. *Journal of Population Economics*, 20(3):643–667, 2007.
- Diana Baumrind. Current patterns of parental authority. *Developmental Psychology*, 4(1, Pt. 2):1–103, 1971. doi: 10.1037/h0030372.
- Roland Bénabou. Workings of a city: Location, education, and production. *The Quarterly Journal of Economics*, 108(3):619–652, 1993. doi: 10.2307/2118403.
- Roland Bénabou and Jean Tirole. Intrinsic and extrinsic motivation. *Review of Economic Studies*, 70(3): 489–520, 2003. doi: 10.1111/1467-937X.00253.
- Roland Bénabou and Jean Tirole. Incentives and prosocial behavior. *American Economic Review*, 96(5): 1652–1678, 2006. doi: 10.1257/aer.96.5.1652.

- Roland Bénabou and Jean Tirole. Identity, morals, and taboos: Beliefs as assets. *Journal of the European Economic Association*, 9(2):285–320, 2011. doi: 10.1111/j.1542-4774.2011.01013.x.
- Marianne Bertrand, Emir Kamenica, and Jessica Pan. Gender identity and relative income within households. *Quarterly Journal of Economics*, 130(2):571–614, 2015.
- A. Bisin and G. Tura. Marriage, fertility, and cultural integration in Italy. Technical Report w26303, National Bureau of Economic Research, 2019.
- A. Bisin and T. Verdier. Beyond the melting pot: cultural transmission, marriage, and the evolution of ethnic and religious traits. *Quarterly Journal of Economics*, 115(3):955–988, 2000.
- A. Bisin and T. Verdier. Agents with imperfect empathy may survive natural selection. *Economics Letters*, 71(2):277–285, 2001a.
- A. Bisin and T. Verdier. On the joint evolution of culture and political institutions: Elites and civil society. *Journal of Political Economy*, 132(5):1485–1564, 2024.
- A. Bisin and T. Verdier. Economic models of cultural transmission. Technical report, National Bureau of Economic Research, 2025.
- A. Bisin, E. Patacchini, T. Verdier, and Y. Zenou. Formation and persistence of oppositional identities. *European Economic Review*, 55(8):1046–1071, 2011.
- Alberto Bisin and Thierry Verdier. The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory*, 97(2):298–319, 2001b.
- Alberto Bisin and Thierry Verdier. The economics of cultural transmission and socialization. In J. Benhabib, A. Bisin, and M. Jackson, editors, *Handbook of Social Economics*, pages 339–416. Elsevier North Holland, 2010.
- Alberto Bisin and Thierry Verdier. Advances in the economic theory of cultural transmission. *Annual Review of Economics*, 15(1):63–89, 2023.
- Alberto Bisin, Eleonora Patacchini, Thierry Verdier, Yves Zenou, et al. *'Bend it Like Beckham': Identity, Socialization and Assimilation*. Centre for Economic Policy Research, 2006.
- Howard Bodenhorn and Christopher S. Ruebeck. The economics of identity and the endogeneity of race. NBER Working Paper 9962, National Bureau of Economic Research, 2003.
- Job Boerma and Loukas Karabarbounis. Reparations and persistent racial wealth gaps. *NBER Macroeconomics Annual*, 37(1):171–221, 2023. doi: 10.1086/723578.
- Giampaolo Bonomi, Nicola Gennaioli, and Guido Tabellini. Identity, beliefs, and political conflict. *Quarterly Journal of Economics*, 2021.
- R. Boyd and P.J. Richerson. *Culture and the Evolutionary Process*. University of Chicago Press, Chicago, 1985.

- Brendan Brundage, Dan J. McGee, and Daniele Tavani. Theoretical approaches in stratification economics. Technical Report 34619, National Bureau of Economic Research, 2026.
- Jean-Paul Carvalho. The paths to narrow identities. *Erasmus Journal for Philosophy and Economics*, 14(2):77–86, 2021.
- L. L. Cavalli-Sforza and M. W. Feldman. *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton University Press, Princeton, 1981.
- Grieve Chelwa, Darrick Hamilton, and James Stewart. Stratification economics: Core constructs and policy implications. *Journal of Economic Literature*, 60(2):377–399, 2022. doi: 10.1257/jel.20211687.
- Robert B. Cialdini and Noah J. Goldstein. Social influence: Compliance and conformity. *Annual Review of Psychology*, 55:591–621, 2004. doi: 10.1146/annurev.psych.55.090902.142015.
- Stephen Cornell and Douglas Hartmann. *Ethnicity and race: Making identities in a changing world*. Sage Publications, 2006.
- Ricardo Dahis, Emily Nix, and Nancy Qian. Choosing racial identity in the united states, 1880–1940. NBER Working Paper 26465, National Bureau of Economic Research, 2019.
- W. Darity, P. Mason, and J. B. Stewart. The economics of identity: the origin and persistence of racial identity norms. *Journal of Economic Behavior and Organization*, 60(3):283–305, 2006.
- William A. Darity. Position and possessions: Stratification economics and intergroup inequality. *Journal of Economic Literature*, 60(2):400–426, 2022. doi: 10.1257/jel.20211690.
- William A. Darity, A. Kirsten Mullen, and Marvin Slaughter. The cumulative costs of racism and the bill for black reparations. *Journal of Economic Perspectives*, 36(2):99–122, 2022. doi: 10.1257/jep.36.2.99.
- William A Darity Jr, Patrick L Mason, and James B Stewart. The economics of identity: the origin and persistence of racial identity norms. *Journal of Economic Behavior & Organization*, 60(3):283–305, 2006.
- Partha Dasgupta and Sanjeev Goyal. Narrow identities. *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft*, pages 395–419, 2019.
- John B. Davis. Stratification economics and identity economics. *Cambridge Journal of Economics*, 39(5):1215–1229, 2015. doi: 10.1093/cje/beu071.
- John B. Davis. Stratification economics as an economics of exclusion. *Journal of Economics, Race, and Policy*, 2(3):163–172, 2019. doi: 10.1007/s41996-019-00026-0.
- Matthias Doepke and Fabrizio Zilibotti. Parenting with style: Altruism and paternalism in intergenerational preference transmission. *Econometrica*, 85(5):1331–1371, 2017.
- Matthias Doepke and Fabrizio Zilibotti. *Love, Money, and Parenting: How Economics Explains the Way We Raise Our Kids*. Princeton University Press, 2019a.

- Matthias Doepke and Fabrizio Zilibotti. Parenting styles and socioeconomic differences in early childhood development. *Annual Review of Economics*, 11:1–26, 2019b.
- W. E. B. Du Bois. *The Souls of Black Folk: Essays and Sketches*. A. C. McClurg & Co., Chicago, IL, 1903. Published April 18, 1903.
- Signithia Fordham and John U Ogbu. Black students’ school success: Coping with the “burden of ‘acting white’”. *The urban review*, 18(3):176–206, 1986.
- Roland G. Fryer, Lisa Kahn, Steven D. Levitt, and Jörg L. Spenkuch. The plight of mixed-race adolescents. *The Review of Economics and Statistics*, 94(3):621–634, August 2012.
- Christian Ghiglini and Andreas Müller. Income redistribution with endogenous social identity. *Available at SSRN 6135378*, 2026.
- Christian Ghiglini, David Juárez-Luna, and Andreas Müller. Class altruism and redistribution. *The Economic Journal*, 131(640):3274–3295, 2021.
- Francesco Giavazzi, Ivan Petkov, and Fabio Schiantarelli. Culture: Persistence and evolution. *Journal of Economic Growth*, 24(2):117–154, 2019.
- Sarah Gillespie, Mirinda M. Morency, Elizabeth Fajemirokun, and Gail M. Ferguson. Promoting identity development, multicultural attitudes, and civic engagement through ethnic studies: Evidence from a natural experiment. *Child Development*, 2025. doi: 10.1111/cdev.14219.
- N. Glazer and D. P. Moynihan. *Beyond the Melting Pot: The Negroes, Puerto Ricans, Jews, Italians, and Irish of New York City*, volume 63. MIT Press, Cambridge, MA, 1970.
- Carolina Gonçalves, Dian Yu, Natasha Keces, and Richard M. Lerner. Within-person fluctuations in ethnic-racial affect and discrimination-based stress: Moderation by average ethnic-racial affect and stress. *Journal of Adolescence*, 97:1173–1185, 2025. doi: 10.1002/jad.12484.
- Milton Myron Gordon. *Assimilation in American life: The role of race, religion, and national origins*. Oxford university press, 1964.
- Gene M. Grossman and Elhanan Helpman. Identity politics and trade policy. *Review of Economic Studies*, 88(3):1101–1134, 2021.
- Robert W. Helsley and Yves Zenou. Social networks and interactions in cities. *Journal of Economic Theory*, 150:426–466, 2014. doi: 10.1016/j.jet.2013.09.009.
- Karla Hoff and Arijit Sen. Homeownership, community interactions, and segregation. *American Economic Review*, 95(4):1167–1189, 2005. doi: 10.1257/0002828054825682.
- Joshua Holm. A model of redistribution under social identification in heterogeneous federations. *Journal of Public Economics*, 143:39–48, 2016. doi: 10.1016/j.jpubeco.2016.08.005.
- Ryo Itoh, Yasuhiro Sato, and Yves Zenou. Intergenerational assimilation of minorities: The role of the majority group. *European Economic Review*, 164:104722, 2024.

- Matthijs Kalmijn and Gerbert Kraaykamp. Determinants of cultural assimilation in the second generation. a longitudinal analysis of values about marriage and sexuality among moroccan and turkish migrants. *Journal of Ethnic and Migration Studies*, 44(5):697–717, 2018.
- Ann H. Kim and Michael J. White. Panethnicity, ethnic diversity, and residential segregation. *American Journal of Sociology*, 115(5):1558–1596, 2010. doi: 10.1086/651375.
- Yn Lê Espiritu. Panethnicity. In *Routledge international handbook of migration studies*, pages 261–271. Routledge, 2019.
- Erik Lindqvist and Robert Östling. Identity and redistribution. *Public Choice*, 155(3):469–491, 2013.
- Daniel E. Martínez and Kelsey E. Gonzalez. Panethnicity as a reactive identity: Primary panethnic identification among latino-hispanics in the united states. *Ethnic and Racial Studies*, 44(4):595–617, 2021. doi: 10.1080/01419870.2020.1752392.
- Moti Michaeli and Jiabin Wu. Fighting polarization with (parental) internalization. *Journal of Economic Behavior & Organization*, 194:124–138, 2022.
- F. M. Moghaddam and E. A. Solliday. Balanced multiculturalism and the challenge of peaceful coexistence in pluralistic societies. *Psychology and Developing Societies*, 3(1):51–72, 1991.
- Michael Omi and Howard Winant. *Racial formation in the United States*. Routledge, 2014.
- Pierre M. Picard and Yves Zenou. Urban spatial structure, employment and social ties. *Journal of Urban Economics*, 104:77–93, 2018. doi: 10.1016/j.jue.2018.01.004.
- Alejandro Portes and Rubén G Rumbaut. *Legacies: The story of the immigrant second generation*. Univ of California Press, 2001.
- Nicholas Sambanis and Moses Shayo. Social identification and ethnic conflict. *American Political Science Review*, 107(2):294–325, 2013.
- Nicholas Sambanis, Stergios Skaperdas, and William C Wohlforth. Nation-building through war. *American Political Science Review*, 109(2):279–296, 2015.
- Alexander Schneeberger and Erin L Krupka. Determinants of norm compliance: Moral similarity and group identification. *Available at SSRN 3969227*, 2021.
- Amartya Sen. *Identity and violence: The illusion of destiny*. WW Norton & Company, 2013.
- Moses Shayo. A model of social identity with an application to political economy: Nation, class, and redistribution. *American Political science review*, 103(2):147–174, 2009.
- Moses Shayo. Social identity and economic policy. *Annual Review of Economics*, 12(1):355–389, 2020.
- Muzafer Sherif. *The Psychology of Social Norms*. Harper, New York, 1936.
- Stergios Skaperdas and Patrick A Testa. National identity, public goods, and modern economic development. *Journal of Comparative Economics*, 53(2):412–432, 2025.

- H. Tajfel. Individuals and groups in social psychology. *British Journal of Social and Clinical Psychology*, 18(2):183–190, 1979.
- H. Tajfel, J. C. Turner, W. G. Austin, and S. Worchel. An integrative theory of intergroup conflict. In *Organizational Identity: A Reader*, pages 56–65. Oxford University Press, 1979.
- Henri Tajfel, editor. *Differentiation Between Social Groups: Studies in the Social Psychology of Intergroup Relations*. Academic Press, London, 1978.
- Henri Tajfel and John C. Turner. An integrative theory of intergroup conflict. In William G. Austin and Stephen Worchel, editors, *The Social Psychology of Intergroup Relations*, pages 33–47. Brooks/Cole, Monterey, CA, 1979.
- Henri Tajfel et al. *Human groups and social categories*, volume 255. Cambridge UK: Cambridge University Press, 1981.
- Emma Thill and Skerdilajda Zana. Economics of cultural change: Openness, interaction, and intergenerational transmission. 2026.
- John C. Turner. Towards a cognitive redefinition of the social group. In Henri Tajfel, editor, *Social Identity and Intergroup Relations*, European Studies in Social Psychology, pages 15–40. Cambridge University Press, Cambridge, 1982.
- Kazuhiro Yuki. Modernization, social identity, and ethnic conflict. *European Economic Review*, 140: 103919, 2021.
- Kazuhiro Yuki. Social identity, redistribution, and development. Technical report, University Library of Munich, Germany, 2023.

APPENDIX 1: Cultural Transmission and Racialism Dynamics

A1) Equilibrium Actions and Utilities:

- 1) Matching II between two individualists: $U^{i,I}(x^h, x^{h'})$

$$x^h = x^{h'} = 0, \text{ and } U^{i,I}(0, 0) = 0$$

- 2) Ingroup matching IR between an individualist h and a racist h' :

$$x^h = 0, x^{h'} = 1, \text{ and } U^{i,IR} = B^i, U^{i,RI} = H^i - C$$

- 3) Ingroup matching RR of two racists:

$$x^h = x^{h'} = 1, \text{ and } U^{i,RR} = B^i + H^i$$

- 4) Outgroup matching IR individualist in group $i \in \{w, b\}$ and racist in group $j \neq i$:

$$x^h = 0, x^{h'} = 1, \text{ and } U^{i,IR} = -D^i, U^{j,RI} = H^j$$

- 5) Outgroup matching RR of two racists $i \in \{w, b\}$ and $j \neq i$:

$$x^h = x^{h'} = 1, U^{i,RR} = -D^i(1 - \delta) + H^i, U^{j,RR} = -D^j(1 - \delta) + H^j$$

B2) Paternalistic Motivations for Individualist and Racist types

Expected paternalistic motivations have to take into account both ingroup and outgroup matchings.

i) Consider an individualist parent I in community i . His expected paternalistic payoffs derived from ingroup matching for an offspring endowed with a individualist trait I or racist trait R write respectively as:

$$\begin{aligned} V^{ii,II} &= q^i B^i \\ V^{ii,IR} &= q^i [B^i - C] + (1 - q^i)(-C) \end{aligned}$$

Similarly, his expected paternalistic payoffs derived from outgroup matching with group $j \neq i$ write as:

$$\begin{aligned} V^{ij,II} &= q^j (-D^i) \\ V^{ij,IR} &= q^j (-D^i(1 - \delta) - C) + (1 - q^j)(-C) \end{aligned}$$

Hence, paternalistic motivations of an individualist parent I under ingroup and outgroup matching are:

$$\begin{aligned} \Delta V^{ii,I} &= V^{ii,II} - V^{ii,IR} = C \\ \Delta V^{ij,I} &= V^{ij,II} - V^{ij,IR} = C - q^j \delta D^i \end{aligned}$$

and the full expected paternalistic motivation taking into account both types of ingroup and outgroup matchings obtains as:

$$\Delta V^{i,I} = Q^{ii} \Delta V^{ii,I} + Q^{ij} \Delta V^{ij,I}$$

Substitution provides that such paternalistic motivation for the individualist trait in community i is frequency dependent and takes the following form:

$$\Delta V^{i,I} \left(q^j_- \right) = Q^{ii} C + Q^{ij} [C - q^j \delta D^i] = C - Q^{ij} q^j \delta D^i$$

ii) Similarly, the expected paternalistic payoffs for a racist parent R in group $i \in \{w, b\}$ derived from ingroup matching, when his offspring is respectively endowed with the racist trait R or the individualist trait I are given by :

$$\begin{aligned} V^{ii,RR} &= q^i B^i - C + (q^i R^i + H^i) \\ V^{ii,RI} &= q^i B^i \end{aligned}$$

As well, the expected paternalistic payoffs derived from outgroup matching with group $j \neq i$ write as:

$$\begin{aligned} V^{ij,RI} &= q^j (-D^i) \\ V^{ij,RR} &= q^j (-D^i(1 - \delta) - C + H^i) + (1 - q^j)(-C + H^i) \end{aligned}$$

Hence, paternalistic motivations of a racist parent R under the two types of matching are given by:

$$\begin{aligned} \Delta V^{ii,R} &= H^i - C + q^i R^i \\ \Delta V^{ij,R} &= H^i - C + q^j \delta D^i \end{aligned}$$

and the expected paternalistic motivation of a racist parent R of group i taking into account both types of ingroup and outgroup matchings obtains as:

$$\Delta V^{i,R} = Q^{ii} \Delta V^{ii,R}(G) + Q^{ij} \Delta V^{ij,R}$$

Substitution provides then paternalistic racist motivation as:

$$\Delta V^{i,R} \left(q^i_+, q^j_+ \right) = H^i - C + Q^{ii} q^i R^i + Q^{ij} q^j \delta D^i \quad (12)$$

QED.

C1) Cultural Dynamics of Racism:

Following (Bisin and Verdier [2001b]), the cultural dynamics of racism in the two groups obtain as:

$$\begin{aligned} \dot{q}^w &= q^w(1 - q^w) [d^{w,R} - d^{w,I}] \\ \dot{q}^b &= q^b(1 - q^b) [d^{b,R} - d^{b,I}] \end{aligned} \quad (13)$$

where $d^{i,R}$ and $d^{i,I}$ are respectively the equilibrium socialization rates of Racist and Individualist parents in group i . Assuming, as usual that parental socialization costs are quadratic $\phi(d) = c \frac{d^2}{2}$, substitution provides:

$$\begin{aligned} \dot{q}^w &= q^w(1 - q^w) c [\Delta V^{w,R}(q^w, q^b)(1 - q^w) - \Delta V^{w,I}(q^b) q^w] \\ \dot{q}^b &= q^b(1 - q^b) c [\Delta V^{b,R}(q^b, q^w)(1 - q^b) - \Delta V^{b,I}(q^w) q^b] \end{aligned} \quad (14)$$

As is clear, the cultural evolution of the racist traits in the two communities are intertwined, and to get a sense of these cultural dynamics, one may consider the phase diagram of this system of differential equations.

The cultural steady state manifold within community $i \in \{w, b\}$, given a frequency q^j of outgroup racials $j \neq i$, is characterized by the following condition:

$$\frac{q^i}{1 - q^i} = \frac{\Delta V^{iR}(q^i, q^j)}{\Delta V^{iI}(q^j)} = \Omega^i(q^i, q^j) = \frac{H^i - C + Q^{ii}q^i R^i + Q^{ij}q^j \delta D^i}{C - Q^{ij}q^j \delta D^i}$$

This can be rewritten as

$$q^j = \Theta^i(q^i) = \frac{C - H^i + H^i q^i - Q^{ii} R^i q^i (1 - q^i)}{Q^{ij} \delta D^i}$$

From assumption 1, $C - H^i < 0$ and $C > \delta D^i \geq Q^{ij} \delta D^i$. Thus $\Theta^i(0) < 0$ and $\Theta^i(1) > 1$. Moreover, $\Theta^i(q^i)$ is a convex quadratic function of q^i having a minimum at $q^i = \frac{1}{2} \left[1 - \frac{H^i}{Q^{ii} R^i} \right]$. Denote $q_L^i = \max \left(0, \frac{1}{2} \left[1 - \frac{H^i}{Q^{ii} R^i} \right] \right)$. Then $\Theta^i(q^i)$ is increasing convex in $q^i \in [q_L^i, 1]$ and there is a unique pair $(q_{\min}^i, q_{\max}^i) \in (q_L^i, 1)^2$ such that $\Theta^i(q_{\min}^i) = 0$ and $\Theta^i(q_{\max}^i) = 1$ and $\Theta^i(q^i)$ is an increasing function mapping $[q_{\min}^i, q_{\max}^i] \subset [0, 1]$ onto $[0, 1]$. Inverting provides the interior cultural steady state manifold $q^i = \tilde{q}^i(q^j) = (\Theta^i)^{-1}(q^j)$ for all $q^j \in [0, 1]$. It is increasing concave in q^j with $\tilde{q}^i(0) = q_{\min}^i$, $\tilde{q}^i(1) = q_{\max}^i$. From this we obtains that there is a unique cultural steady state within group :

$$q^i = \tilde{q}^i(q^j) \in (0, 1) \text{ for } i, j \in \{b, w\} \text{ and } i \neq j$$

It is a simple matter to see that the two curves $q^w = \tilde{q}^w(q^b)$ and $q^b = \tilde{q}^b(q^w)$ are concave increasing and illustrated in figure (1) in the main text. **QED.**

D1) Existence and Uniqueness of Interior Cultural Steady State (q_*^w, q_*^b)

An interior cultural steady state $(q_*^w, q_*^b) \in (0, 1)^2$ of the joint cultural dynamics across groups is determined by the following equations:

$$q_*^w = \tilde{q}^w(q_*^b) \text{ and } q_*^b = \tilde{q}^b(q_*^w)$$

Defining the function $\Upsilon(x) = [\tilde{q}^w \circ \tilde{q}^b](x) - x$, We have $\Upsilon(0) = \tilde{q}^w(q_{\min}^b) > \tilde{q}^w(0) = q_{\min}^w > 0$. $\Upsilon(1) = \tilde{q}^w(q_{\max}^b) - 1 < \tilde{q}^w(1) - 1 < 0$. Moreover $\Upsilon(x)$ is such that $\Upsilon'(x) = (\tilde{q}^w)'(\tilde{q}^b(x)) \cdot (\tilde{q}^b)'(x) - 1$ and $\Upsilon''(x) = (\tilde{q}^w)''(\tilde{q}^b(x)) \cdot [(\tilde{q}^b)'(x)]^2 + (\tilde{q}^w)'(\tilde{q}^b(x)) \cdot (\tilde{q}^b)''(x) < 0$ as $\tilde{q}^w(\cdot)$ and $\tilde{q}^b(\cdot)$ are concave functions. Thus $\Upsilon(x)$ is a concave continuous function such that function $\Upsilon(0) > 0$ and $\Upsilon(1) < 0$. It should have a unique point x^* such that $\Upsilon(x^*) = 0$. From this, we recover the unique interior steady state (q_*^w, q_*^b) given by $q_*^w = [\tilde{q}^w \circ \tilde{q}^b](x^*)$ and $q_*^b = \tilde{q}^b(q_*^w)$. **QED**

E1) Effect of Social Integration on the Cultural Dynamics of Racism:

- **Comparative statics of s on steady state manifold $\tilde{q}^i(q^j)$:** For $i, j \in \{b, w\}$ and $i \neq j$, there is a unique value $\bar{q}^j \in [0, 1]$ such that $\frac{\partial \tilde{q}^i}{\partial s} \leq 0$ if and only if $q^j \leq \bar{q}^j$.

Proof: Recall that $\tilde{q}^i(q^j)$ is determined by

$$\frac{q^i}{1 - q^i} = \Omega^i(q^i, q^j) \quad (15)$$

with

$$\Omega^i(q^i, q^j) = \frac{H^i + Q^{ii}q^i R^i}{C - Q^{ij}q^j \delta D^i} - 1$$

Therefore the sign of $\frac{\partial \tilde{q}^i}{\partial s}$ is the same as the sign of $\frac{\partial \Omega^i}{\partial s}$ evaluated at $q^i = \tilde{q}^i(q^j)$ for $q^j \in [0, 1]$. Simple differentiation provides that

$$\text{sign} \left[\frac{\partial \Omega^i(q^i, q^j)}{\partial s} \right] = \text{sign} [-q^i R^i C + (H^i + q^i R^i) q^j \delta D^i]$$

Thus $\text{sign} \left[\frac{\partial \Omega^i(q^i, q^j)}{\partial s} \right] < 0$ if and only if

$$H^i \frac{q^j \delta D^i}{C - q^j \delta D^i} < q^i R^i \quad (16)$$

Denote for convenience

$$\begin{aligned} Z^i &= Q^{ij} = (1 - \lambda^i)s \in [0, 1], \quad A^i = 1 - Z^i = Q^{ii} = 1 - (1 - \lambda^i)s \in [0, 1] \\ d^i &= \delta D^i \end{aligned}$$

Recall as well the parameter restrictions from assumption (1):

$$0 < d^i < C < H^i,$$

so that $C - q^j d^i > 0$ for all $q^j \in [0, 1]$. (15) rewrites as

$$\frac{1}{1 - q^i} = \frac{H^i + A^i q^i R^i}{C - Z^i q^j d^i}, \quad (17)$$

given that (16) rewrites as

$$\frac{H^i}{R^i} \frac{q^j d^i}{C - q^j d^i} < q^i \quad (18)$$

Combining (15) with (18), it follows that $\text{sign} \left[\frac{\partial \Omega^i(q^i, q^j)}{\partial s} \right] < 0$ if and only if

$$(1) \quad \frac{1}{1 - q^i} = \frac{H^i + A^i q^i R^i}{C - Z^i q^j d^i}, \quad (2) \quad \frac{H^i}{R^i} \frac{q^j d^i}{C - q^j d^i} < q^i.$$

(17) provides

$$q_j d^i = \frac{C - (1 - q^i)(H^i + A^i q^i R^i)}{Z^i}$$

and

$$C - q_j d^i = C - \frac{C - (1 - q^i)(H^i + A^i q^i R^i)}{Z^i} = \frac{(1 - q^i)(H^i + A^i q^i R^i) - A^i C}{Z^i},$$

because $Z^i - 1 = -A^i$.

Thus

$$\frac{q^j d^i}{C - q^j d^i} = \frac{C - (1 - q^i)(H^i + A^i q^i R^i)}{(1 - q^i)(H^i + A^i q^i R^i) - A^i C}.$$

And (18) rewrites as

$$\frac{H^i}{R^i} \frac{C - (1 - q^i)(H^i + A^i q^i R^i)}{(1 - q^i)(H^i + A^i q^i R^i) - A^i C} < q^i.$$

Since $R^i > 0$ and $(C - q_j d^i) > 0$, this inequality is equivalent to:

$$H^i \left(C - (1 - q^i)(H^i + A^i q^i R^i) \right) < q^i R^i \left((1 - q^i)(H^i + A^i q^i R^i) - A^i C \right).$$

Rearranging terms, this rewrites as:

$$C(H^i + A^i q^i R^i) < (1 - q^i)(H^i + A^i q^i R^i)(H^i + q^i R^i).$$

Because $H^i + A^i q^i R^i > 0$, this last condition is equivalent to :

$$(1 - q^i)(H^i + q^i R^i) - C > 0.$$

or finally

$$\sigma(q^i) = R^i(q^i)^2 + (H^i - R^i)q^i + (C - H^i) < 0.$$

as $C < H^i$, the quadratic function $\sigma(q^i)$ is negative at $q^i = 0$ and positive at $q^i = 1$, there is a unique root $\bar{q}^i \in (0, 1)$ such that $\sigma(\bar{q}^i) = 0$, and therefore

$$\sigma(q^i) < 0 \iff q^i < \bar{q}^i \quad .$$

with

$$\bar{q}^i = \frac{R^i - H^i + \sqrt{(H^i - R^i)^2 + 4R^i(H^i - C)}}{2R^i}$$

Define then $\hat{q}^j$ defined by

$$\begin{aligned} \hat{q}^j &= \frac{1}{d^i} \frac{C - (1 - \bar{q}^i)(H^i + A^i \bar{q}^i R^i)}{B^i} \\ &= \frac{R^i \bar{q}^i (1 - \bar{q}^i)}{d^i} \end{aligned}$$

Define then

$$\bar{q}^j = \min [\hat{q}^j, 1] = \min \left[\frac{R^i \bar{q}^i (1 - \bar{q}^i)}{\delta D^i}, 1 \right]$$

Along the manifold curve $\tilde{q}^i(q^j)$, condition (18) holds if and only if

$$q^j < \bar{q}^j$$

(with the understanding that when $\bar{q}^j = 1$ then ((18)) holds for all admissible q^j). From this, we deduce that $\frac{\partial \tilde{q}^i}{\partial s} \leq 0$ if and only if $q^j \leq \bar{q}^j$. **QED**

APPENDIX 2: A General Set-up of Identity Adoption and Cultural Transmission

In this appendix, we present a general approach of identity formation within generations as developed by Shayo (2009), embedded into a model of intergenerational cultural transmission. Individuals with fixed individual characteristics, choose a group identity and then take individual private actions. Choices reflect a basic trade-off between how the individual 'fits' within the group - that is, how his/her characteristics are close enough to the group's prototype - and the social esteem or social status that belonging to that group confers - which itself may depend on social norms and social prescriptions at the level of the group, and on in-group and out-groups actions.

Specifically, let a large population with mass normalized to 1, made of individuals, each defined by a fixed type $i \in I$ for I a finite set of types. The distribution of characteristic types i in the population for the moment is fixed and represented by a frequency distribution $\mathbf{Q} = (q^i)_{i \in I}$ on I that belongs to the simplex $\Delta(I) = \left\{ (q^i)_{i \in I} \mid \sum_{i \in I} q^i = 1 \right\}$.

An individual with attributes i takes a social action $x^i \in \mathcal{X}$ generating an individual payoff $\pi^i(x^i, X)$, where X is an aggregator of individual actions in society : $X = \Theta \left((x^i)_{i \in I}, \mathbf{Q} \right)$, that depends on the full profile of actions $(x^i)_{i \in I}$, and the distribution of types $\mathbf{Q}$.³⁷

Individuals in group i choose to identify to a social group $g \in \mathcal{G} = \{1, \dots, G\}$, where g represents the archetype vector of attributes of the group. Identifying to g for an individual i induces a psychological cost related to the social distance between the groups' characteristics $\delta(i, g)$. At the same time, identification confers some social esteem related to the social status of the group $s^g = s(x^i, \pi^g, x^g, q^g)$, where π^g, x^g, q^g denote, respectively, the group average profits, actions, and relative size of the group g . The dependence of s^g on own action x^i may reflect the existence of social norms that regulate individual behaviors, providing positive or negative social recognition.

The utility function of an individual with attributes i who identifies to g is

$$U^i(x^i, X; g) = \pi^i(x^i, X) - \delta(i, g) + \alpha s(x^i, \pi^g, x^g, q^g);$$

where $\alpha > 0$ reflects the relative importance of social status with respect to cognitive distance on the individual's preferences.³⁸

Individuals with attributes i choose first the group(s) they identify to g ; and in a second stage they choose their optimal action x^i . In both stages individual take for given the map $G(\cdot)$ that associates to each type i his identity group $g = G(i)$ and the second state choices of all individuals, x^j , for all characteristics j , not just the individual's own.³⁹

By backward induction, consider the choice of action of an individuals' i who identifies to g and takes X as given:

$$\begin{aligned} & \max_{x^i} U^i(x^i, X; g) \\ \text{s.t. } & \pi^g = \sum_{g=G(j)} \pi^j, \quad x^g = \sum_{g=G(j)} x^j, \quad q^g = \sum_{g=G(j)} q^j \end{aligned}$$

³⁷In the main text, we simply consider the average action in society $X = \sum_{i \in I} q^i x^i$.

³⁸Here, with no public policy or voting, identity reshapes *private* utility via conformity and group-referenced motives, and group status feeds back from realized private outcomes.

³⁹Formally, we are formulating an atomistic game, where the population is composed of a mass of individuals for any characteristic i .

Under regularity conditions, one obtains an optimal behavior $\bar{x}^i = \bar{x}^i(X, \pi^g, x^g, q^g, g)$. At equilibrium, one must have the fixed point condition:⁴⁰

$$X = \Theta \left((x^i)_{i \in I}, \mathbf{Q} \right),$$

$$x^i = \bar{x}^i(X, \pi^g, x^g, q^g, g) \text{ for all } i \in I \text{ and } G(i) = g$$

$$\pi^g = \sum_{j : G(j)=g} \pi^j(x^j, X), \quad x^g = \sum_{j : G(j)=g} x^j, \quad q^g = \sum_{j : G(j)=g} q^j$$

inducing - with enough regularity conditions - individual equilibrium choices $x^i(G, \mathbf{Q})$ and aggregate index $X(G, \mathbf{Q})$. Associated to this equilibrium, one derives equilibrium utility payoffs:

$$V^i(G, \mathbf{Q}, g) = U^i(x^i(G, \mathbf{Q}), X(G, \mathbf{Q}), g)$$

The identification choice of individuals with characteristics i is then written as:

$$\max_g V^i(G, \mathbf{Q}, g),$$

and it induces a map $g = F(i, G, \mathbf{Q})$. At equilibrium

$$G(i) = F(i, G, \mathbf{Q}), \text{ for all } i \in I \quad (19)$$

Conceptually, this provides an equilibrium pattern of identity choices $G^* = G^*(\mathbf{Q})$.⁴¹ Associated to this, one gets the equilibrium payoffs:

$$W^i(\mathbf{Q}) = V^i(G^*(\mathbf{Q}), \mathbf{Q}, G^*(i)) \text{ for all } i \in I$$

One may embed the previous intragenerational set-up into a dynamic model of intergenerational cultural evolution, that shape the distribution of types $\mathbf{Q}$ by evolutionary forces as encapsulated in cultural replicator dynamics. For instance, taking the continuous time approximation, the evolution of the population distribution $\mathbf{Q}(t) = (q^i(t))_{i \in I}$ can be governed by a set of differential equations:

$$\dot{q}^i(t) = q^i(t) \cdot (F^i(\mathbf{Q}(t)) - \bar{F}(\mathbf{Q}(t))) \text{ and } \mathbf{Q}(t) = (q^i(t))_{i \in I} \quad (20)$$

where $F^i(\mathbf{Q}(t))$ is the cultural fitness of attributes i list and $\bar{F}(\mathbf{Q}(t))$ is the average cultural fitness in the population of attributes $i \in I$ in generation t .

In the canonical Bisin and Verdier [2001a] set -up, a precise parametrization of the cultural fitness functions is obtained. To see that, consider a simple generalization of (1) to a multi-trait population with the following transition probabilities that a child from a family with trait i is socialized to trait j :

$$P^{ii}(t) = d^i(t) + (1 - d^i(t))q^i(t), \quad P^{ij}(t) = (1 - d^i(t))q^j(t) \quad (21)$$

⁴⁰In the main text, we abstract from the dependence of s^g on individual actions x^i . This simplifies the presentation of the fixed point characterizing the equilibrium actions $x^i(G, \mathbf{Q})$. Indeed, in such a case as individuals are atomistic, the optimal actions $\bar{x}^i(X, \pi^g, x^g, q^g, g)$ only depend on the aggregate index X , and not on the (endogenous) arguments π^g, x^g .

⁴¹Note that even if there is a well-defined unique second stage equilibrium outcome $(x^i(G, \mathbf{Q}))_{i \in I}$ for each identity group configuration $G(\cdot)$, one may still have multiple identity equilibria $G^*(\mathbf{Q})$ satisfying (19). In such a case, the cultural dynamics described afterwards are conditional on a specific equilibrium selection process.

for all $j \neq i$ and $(i, j) \in I^2$. With such a process of cultural transmission, cultural dynamics are easily obtained as:

$$\dot{q}^i(t) = q^i(t) \left[d^i(t) - \sum_{j \in I} q^j(t) d^j(t) \right] \text{ for all } i \in I \quad (22)$$

Connecting (20) and (22) then shows that the cultural fitness of type i is then simply $F^i(\mathbf{Q}(t)) = d^i(t) = d^i(\mathbf{Q}(t))$ with $d^c(\mathbf{Q}(t))$ being the optimal vertical rate of socialization of a parent of type i :

$$d^i(\mathbf{Q}) = \arg \max_d \left[\begin{array}{l} [d + (1-d)q^i] \cdot W^i(\mathbf{Q}) + \\ +(1-d) \sum_{\substack{j \neq i \\ j \in I}} q^j V^{ij}(\mathbf{Q}) - C(d) \end{array} \right]$$

In this last equation

$$V^{ij}(\mathbf{Q}) = U^i(x^j(G^*, \mathbf{Q}), X(G^*, \mathbf{Q}), G^*(i))$$

is the utility payoff of the child adopting trait j as perceived by parent of type i with identity $G^*(i)$, understanding that such a child will consequently pick an equilibrium identity $G^*(j)$ and undertake an individual optimal action $x^j(G^*(\mathbf{Q}), \mathbf{Q})$.⁴² Looking at the first order condition provides:

$$W^i(\mathbf{Q})(1 - q^i) - \sum_{\substack{j \neq i \\ j \in I}} q^j V^{ij}(\mathbf{Q}) = C'(d^i)$$

which can be rewritten as

$$\sum_{j \in I} q^j \Delta W^{ij}(\mathbf{Q}) = C'(d^i)$$

with

$$\Delta W^{ij}(\mathbf{Q}) = W^i(\mathbf{Q}) - V^{ij}(\mathbf{Q})$$

describing the "paternalistic" value for a parent of type i of having a child of his own type rather than the alternative type j . One may then denote $\Delta \mathbf{W}^i(\mathbf{Q})$ the vector of paternalistic values of cultural type i over all possible alternative traits j : $\Delta \mathbf{W}^i(\mathbf{Q}) = [\Delta W^{ij}(\mathbf{Q})]_{j \in I}$ and $\Delta \mathbf{W}(\mathbf{Q})$ the corresponding matrix $[\Delta W^{ij}(\mathbf{Q})]_{(i,j) \in I^2}$ of paternalistic cultural values, where each line is given by the vector line $\Delta \mathbf{W}^i(\mathbf{Q})^\top$. Then the optimal vertical socialization rate of type i $d^i = d^i(\mathbf{Q})$ obtains as:

$$C'(d^i) = \Delta \mathbf{W}^i(\mathbf{Q})^\top \cdot \mathbf{Q}$$

The cultural fitness of type i writes as:

$$F^i(\mathbf{Q}(t)) = d^i(\mathbf{Q}(t)) = C'^{-1} \left[\Delta \mathbf{W}^i(\mathbf{Q}(t))^\top \cdot \mathbf{Q}(t) \right]$$

while average cultural fitness is given by:

$$\bar{F}(\mathbf{Q}(t)) = \sum_{j \in I} q^j(t) \cdot d^j(\mathbf{Q}(t)) = \sum_{j \in I} q^j(t) \cdot C'^{-1} \left[\Delta \mathbf{W}^j(\mathbf{Q}(t))^\top \cdot \mathbf{Q}(t) \right]$$

As can be immediately seen, equilibrium patterns of identity choices $G^*(\mathbf{Q}_t)$ depend on the distribution of longer term cultural traits. Conversely, cultural fitness functions $F^i(\mathbf{Q}(t))$ depend, through

⁴²Note that by extension $W^i(\mathbf{Q}) = V^{ii}(\mathbf{Q})$

the paternalistic value matrix $\Delta \mathbf{W}(\mathbf{Q}(t))$, on the shape of equilibrium intragenerational identity choices, $\mathbf{G}^*(.)$. Consequently, the dynamic path of cultural evolution $[\mathbf{Q}(t)]_{t \in [0, \infty[}$ depends reciprocally on intra-generational identity choices. Both inter-generational cultural transmission and intra-generational identity choices are intertwined.

APPENDIX 3: Broad Versus Narrow identities

A3) Equilibrium pattern of identity choices

The description of the equilibrium pattern of identity choices depends on the comparison of the two thresholds $\hat{q}^a$ and $1 - \hat{q}^b$. It is simple to see that $\hat{q}^a > 1 - \hat{q}^b$ if and only if $\gamma/\beta < \frac{1}{2R^2} \frac{\theta^a \theta^b}{\theta^a + \theta^b}$. From this and conditions (9) one immediately obtains that:

- **Result A3 :** *i) If $\gamma/\beta < \frac{1}{2R^2} \frac{\theta^a \theta^b}{\theta^a + \theta^b}$, there is a full "narrow" social identity structure associated to the two cultural traits when the cultural population is "balanced" (ie. q taking intermediate values in an interval $[1 - \hat{q}^b, \hat{q}^a]$). ii) If $\gamma/\beta \geq \frac{1}{2R^2} \frac{\theta^a \theta^b}{\theta^a + \theta^b}$, there is a full "broad" social identity structure associated to the two cultural traits when the cultural population is "balanced" (ie. q taking intermediate values in an interval $[\hat{q}^a, 1 - \hat{q}^b]$)*

B3) Paternalistic motivations of parents

The paternalistic utility of a parent of type i with identity g to have a child of type j writes as:

$$\begin{aligned} V_g^{ij} &= \pi^i(x^j) - \beta(\delta(i, g)) + \gamma s^g \\ &= R \cdot x^j - \theta^i \frac{(x^j)^2}{2} - \beta(\delta(i, g)) + \gamma s^g \end{aligned}$$

From this, we can write the paternalistic motivations of a parent of type i with a narrow identity $g = \{i\}$ to transmit his trait i rather than the other trait j as:

$$\begin{aligned} \Delta V_i^i &= V_i^{ij} - V_i^{ij} \\ &= \left(R \cdot x^i - \theta^i \frac{(x^i)^2}{2} \right) - \left(R \cdot x^j - \theta^i \frac{(x^j)^2}{2} \right) - \beta[\delta(i, i) - \delta(j, i)] \\ &= \frac{R^2}{2\theta^j} \frac{[\theta^j - \theta^i]^2}{\theta^i \theta^j} + \beta \end{aligned}$$

Conversely, for a parent with trait i and a broad identity ab , the paternalistic motivations to transmit trait i rather than trait j writes as have:

$$\begin{aligned} \Delta V_{ab}^i &= V_{ab}^{ij} - V_{ab}^{ij} \\ &= \left(R \cdot x^i - \theta^i \frac{(x^i)^2}{2} \right) - \left(R \cdot x^j - \theta^i \frac{(x^j)^2}{2} \right) - \beta[\delta(i, ab) - \delta(j, ab)] \\ &= \frac{R^2}{2\theta^j} \frac{[\theta^j - \theta^i]^2}{\theta^i \theta^j} + \beta[2q^i - 1] \end{aligned}$$

C3) Equilibrium Socialization efforts

For convenience denote for $i \in \{a, b\}$:

$$\Delta V^i = \frac{R^2}{2\theta^j} \frac{[\theta^j - \theta^i]^2}{\theta^i \theta^j} \quad (23)$$

The equilibrium rates of cultural transmission of both types of parents write as a function of $q = q_a = 1 - q_b$:

$$d^a(q) = \begin{cases} d_N^a(q) = c[\Delta V^a + \beta](1 - q) & \text{when } q < \hat{q}^a \\ d_B^a(q) = c[\Delta V^a + \beta(2q - 1)](1 - q) & \text{when } q \geq \hat{q}^a \end{cases} \quad (24)$$

and

$$d^b(q) = \begin{cases} d_N^b(q) = c[\Delta V^b + \beta]q & \text{when } q > 1 - \hat{q}^b \\ d_B^b(q) = c[\Delta V^b + \beta(1 - 2q)]q & \text{when } q \leq 1 - \hat{q}^b \end{cases} \quad (25)$$

Consider the following assumption:

$$\text{Assumption H: } 3\beta < \Delta V^a = \frac{R^2}{2\theta^b} \frac{[\theta^b - \theta^a]^2}{\theta^a \theta^b} \text{ and } 3\beta < \Delta V^b = \frac{R^2}{2\theta^a} \frac{[\theta^b - \theta^a]^2}{\theta^a \theta^b}$$

Simple inspection shows that, under assumption **H**, $d^a(q)$ and $d^b(q)$ are both positive. Moreover, $d^a(q)$ is decreasing in q while $d^b(q)$ is increasing in q , implying that they follow the cultural substitutability property of (Bisin and Verdier [2001b]).

Finally the following assumption **H'** ensures that $d^a(q)$ and $d^b(q)$ are both below 1:

$$\text{Assumption H': } c \leq \min \left\{ \frac{1}{[\Delta V^a + \beta]}, \frac{1}{[\Delta V^b + \beta]} \right\}$$

D3) Cultural dynamics in each identity regime

i) In a "full narrow identity" regime: the cultural dynamics converge towards an interior steady q^N characterized by the condition $d_N^a(q^N) = d_N^b(q^N)$. Substitution provides:

$$q^N = \frac{\Delta V^a + \beta}{\Delta V^a + \Delta V^b + 2\beta} = \frac{R^2 \frac{\theta^a}{\theta^a + \theta^b} + \beta \frac{2[\theta^a \theta^b]^2}{[\theta^b - \theta^a]^2(\theta^a + \theta^b)}}{R^2 + 2\beta \frac{2[\theta^a \theta^b]^2}{[\theta^b - \theta^a]^2(\theta^a + \theta^b)}} \quad (26)$$

ii) In a "Full Broad Identity": the dynamics converge towards an interior steady q^B characterized by the condition $d_B^a(q^B) = d_B^b(q^B)$, thus after substitution:

$$q^B = \frac{\Delta V^a - \beta}{\Delta V^a + \Delta V^b - 2\beta} = \frac{R^2 \frac{\theta^a}{\theta^a + \theta^b} - \beta \frac{2[\theta^a \theta^b]^2}{[\theta^b - \theta^a]^2(\theta^a + \theta^b)}}{R^2 - 2\beta \frac{2[\theta^a \theta^b]^2}{[\theta^b - \theta^a]^2(\theta^a + \theta^b)}} \quad (27)$$

iii) In a "Mixed regime" with "narrow identity" for trait a and "broad identity" for trait b , the cultural dynamics converge towards an interior steady q^{Mb} characterized by the condition $d_N^a(q^{Mb}) = d_B^b(q^{Mb})$, thus q^{Mb} is implicitly given by the condition:

$$\frac{q^{Mb}}{1 - q^{Mb}} = \frac{\Delta V^a + \beta}{\Delta V^b + \beta(1 - 2q^{Mb})}$$

or after substitution,

$$q^{Mb} = \frac{\Delta V^a + \beta}{\Delta V^a + \Delta V^b + 2\beta(1 - q^{Mb})} = \frac{R^2 \frac{\theta^a}{\theta^a + \theta^b} + \beta \frac{2[\theta^a \theta^b]^2}{[\theta^b - \theta^a]^2 (\theta^a + \theta^b)}}{R^2 + 2\beta(1 - q^{Mb}) \frac{2[\theta^a \theta^b]^2}{[\theta^b - \theta^a]^2 (\theta^a + \theta^b)}} \quad (28)$$

iv) Finally in a "Mixed regime" with "narrow identity" for trait b and "broad identity" for trait a , the cultural dynamics converge towards the interior steady q^{Ma} characterized by the condition $d_B^a(q^{Ma}) = d_N^b(q^{Ma})$. Thus it is implicitly determined by the condition:

$$\frac{q^{Ma}}{1 - q^{Ma}} = \frac{\Delta V^a + \beta(2q^{Ma} - 1)}{\Delta V^b + \beta}$$

or after substitution,

$$q^{Ma} = \frac{\Delta V^a + \beta(2q^{Ma} - 1)}{\Delta V^a + \Delta V^b + 2\beta q^{Ma}} = \frac{R^2 \frac{\theta^a}{\theta^a + \theta^b} + \beta(2q^{Ma} - 1) \frac{2[\theta^a \theta^b]^2}{[\theta^b - \theta^a]^2 (\theta^a + \theta^b)}}{R^2 + 2\beta q^{Ma} \frac{2[\theta^a \theta^b]^2}{[\theta^b - \theta^a]^2 (\theta^a + \theta^b)}} \quad (29)$$

E3) Comparison between cultural steady states

First given that $\Delta V^a < \Delta V^b$ it follows that the function $\phi(x) = \frac{\Delta V^a + x}{\Delta V^a + \Delta V^b + 2x}$ is increasing in x . Thus

$$q^N = \phi(\beta) > \phi(-\beta) = q^B \quad (30)$$

Second, the function $\psi(q) = \frac{\Delta V^a + \beta}{\Delta V^a + \Delta V^b + 2\beta(1 - q)}$ is increasing in q and such that $\psi(0) = \frac{\Delta V^a + \beta}{\Delta V^a + \Delta V^b + 2\beta} = q^N$. From this it follows that

$$q^{Mb} = \psi(q^{Mb}) > \psi(0) = q^N \quad (31)$$

Finally note that q^B is determined by the relationship

$$\frac{q}{1 - q} = \rho(q) = \frac{\Delta V^a + \beta(2q - 1)}{\Delta V^b + \beta(1 - 2q)} \quad (32)$$

while q^{Ma} is determined by the relationship

$$\frac{q}{1 - q} = \Lambda(q) = \frac{\Delta V^a + \beta(2q - 1)}{\Delta V^b + \beta} \quad (33)$$

Given that $\Lambda(q) \leq \rho(q)$ for all $q \in (0, 1]$, it follows that $q^{Ma} < q^B$.

From the previous discussion, we conclude that:

$$q^{Ma} < q^B < q^N < q^{Mb} \quad (34)$$

F3) Cultural dynamics under endogenous identity choices

We have the following result characterizing the steady state cultural population and identity choice regime:

- **Result F2:** Assume that assumption H holds and that $\beta < (\Delta V^b - \Delta V^a)$. Then there exists $\gamma_0, \gamma_1, \gamma_2, \gamma_3 > 0$ such that in the long run : i) if $\gamma < \gamma_0$, a "Full Narrow Identity" regime prevails and $q^* = q^N$; ii) if $\gamma \in (\gamma_0, \gamma_1)$, a "Mixed Identity" regime prevails with trait a associated to "narrow" identity, and trait b indifferent between "narrow" or "broad" identity, and $q^* = 1 - \hat{q}_B(\gamma)$; iii) if $\gamma \in (\gamma_1, \gamma_2)$, a "Mixed Identity" regime prevails with trait a associated to "narrow" identity, and trait b associated to "broad" identity, and $q^* = q^{Mb}$; iv) if $\gamma \in (\gamma_2, \gamma_3)$, a "Mixed Identity" regime prevails with trait a indifferent between "narrow" and "broad" identity, and trait b associated to "broad" identity, and $q^* = \hat{q}_A(\gamma)$; v) If $\gamma \geq \gamma_3$, a "Full Broad Identity" regime prevails and $q^* = q^B$.

Proof: Recall that $\hat{q}^a(\gamma) = 1 - \frac{2\gamma}{\beta} \frac{R^2}{\theta^b}$, $1 - \hat{q}^b(\gamma) = \frac{2\gamma}{\beta} \frac{R^2}{\theta^a}$. Denote $\hat{\gamma}$ the value of γ such that $\hat{q}^a(\hat{\gamma}) = 1 - \hat{q}^b(\hat{\gamma})$. We get $\hat{\gamma} = \frac{\beta}{2R^2} \frac{\theta^a \theta^b}{\theta^a + \theta^b}$, also simple substitution provides $1 - \hat{q}^b(\hat{\gamma}) = \hat{q}^a(\hat{\gamma}) = \frac{\theta^b}{\theta^a + \theta^b} = \bar{q} > \frac{1}{2}$.

First, it is easy to see for $\beta < (\Delta V^b - \Delta V^a)$, one has necessarily $q^{Mb} < \bar{q}$. Indeed q^{Mb} is determined by the condition

$$q^{Mb} = \frac{\Delta V^a + \beta}{\Delta V^a + \Delta V^b + 2\beta(1 - q^{Mb})}$$

This is less than $\bar{q} = \frac{\Delta V^b}{\Delta V^a + \Delta V^b}$ when

$$\frac{\Delta V^a + \beta}{\Delta V^a + \Delta V^b + 2\beta(1 - q)} < \frac{\Delta V^b}{\Delta V^a + \Delta V^b}$$

for all $q \in [0, 1]$. A sufficient condition for this is then $\beta < (\Delta V^b - \Delta V^a)$.

Given this, and from the fact that $1 - \hat{q}^b(\gamma)$ is increasing in γ and $\hat{q}^a(\gamma)$ is decreasing in γ , and the fact that $q^B < q^N < q^{Mb}$, one immediately obtain that there exists $\gamma_0, \gamma_1, \gamma_2, \gamma_3$ such that $0 < \gamma_0 < \gamma_1 < \hat{\gamma} < \gamma_2 < \gamma_3$ and $1 - \hat{q}^b(\gamma_0) = q^N$, $1 - \hat{q}^b(\gamma_1) = q^{Mb}$, $\hat{q}^a(\gamma_2) = q^{Mb}$ and $\hat{q}^a(\gamma_3) = q^B$.

i) case $\gamma < \gamma_0$. In such a case, $1 - \hat{q}^b(\gamma) < q^N < \hat{q}^a(\gamma)$. Simple inspection provides that:

- for all $q_t < 1 - \hat{q}^b(\gamma) < q^N$, $d^b(q_t) = d_B^b(q_t) < d_N^b(q_t) < d_N^a(q_t) = d^a(q_t)$ thus $q_{t+1} > q_t$.
- for $1 - \hat{q}^b(\gamma) < q_t < q^N < \hat{q}^a(\gamma)$, $d^b(q_t) = d_N^b(q_t) < d_N^a(q_t) = d^a(q_t)$ thus $q_{t+1} > q_t$.
- for $q^N < q_t < \hat{q}^a(\gamma)$, $d^b(q_t) = d_N^b(q_t) > d_N^a(q_t) = d^a(q_t)$ thus $q_{t+1} < q_t$.
- for $\hat{q}^a(\gamma) \leq q_t$ one has $d^b(q_t) = d_N^b(q_t) > d_N^a(q_t) > d_B^a(q_t) = d^a(q_t)$ thus $q_{t+1} < q_t$.

Consequently the cultural dynamics converge towards $q^* = q^N$, and in such a situation, there is a "Full Narrow Identity" regime as $1 - \hat{q}^b(\gamma) < q^N < \hat{q}^a(\gamma)$.

ii) case $\gamma_0 < \gamma < \gamma_1$. in such a case we have $q^N < 1 - \hat{q}^b(\gamma) < q^{Mb}$ and given that $q^{Mb} < \bar{q}$, one also has $\gamma_1 < \hat{\gamma}$. Consequently $q^{Mb} < \hat{q}^a(\gamma)$. From this it follows that

- for $q_t < q^N < 1 - \hat{q}^b(\gamma)$, $d^b(q_t) = d_B^b(q_t) < d_N^b(q_t) < d_N^a(q_t) = d^a(q_t)$; thus $q_{t+1} > q_t$.
- for $q^N < q_t \leq 1 - \hat{q}^b(\gamma)$, $d^b(q_t) = d_B^b(q_t) < d_N^a(q_t) = d^a(q_t)$ given that $q_t < 1 - \hat{q}^b(\gamma) < q^{Mb}$; thus $q_{t+1} > q_t$.
- for $1 - \hat{q}^b(\gamma) < q_t$, $d^b(q_t) = d_N^b(q_t) > d_N^a(q_t) > d_B^a(q_t) = d^a(q_t)$ as $q^N < 1 - \hat{q}^b(\gamma) < q_t$. Thus $d^b(q_t) > d^a(q_t)$ and $q_{t+1} < q_t$.

From this we conclude that the cultural dynamics should converge towards $1 - \hat{q}^b(\gamma)$. In such a situation, there is a "Mixed Identity" regime with cultural trait b indifferently associated to a "narrow" or "broad" identity", and trait a associated to a "narrow" identity.

iii) case $\gamma_1 < \gamma < \gamma_2$. In such a case, we have $q^{Mb} < 1 - \hat{q}^b(\gamma)$ and given that $q^{Mb} < \bar{q}$, one also has $\gamma_2 > \hat{\gamma}$. Consequently $q^{Mb} < \hat{q}^a(\gamma)$.

a) A first situation is when $\gamma_1 < \gamma \leq \hat{\gamma}$. In such a case $q^{Mb} < 1 - \hat{q}^b(\gamma) \leq \hat{q}^a(\gamma)$. Therefore

- for $q_t < q^{Mb}$, $d^b(q_t) = d_B^b(q_t) < d_N^a(q_t) = d^a(q_t)$; thus $q_{t+1} > q_t$.
- for $q^{Mb} \leq q_t \leq 1 - \hat{q}^b(\gamma)$, $d^b(q_t) = d_B^b(q_t) > d_N^a(q_t) = d^a(q_t)$, as $1 - \hat{q}^b(\gamma) < \hat{q}^a(\gamma)$; thus $q_{t+1} < q_t$.
- for $1 - \hat{q}^b(\gamma) < q_t$, $d^b(q_t) = d_B^b(q_t) > d_N^a(q_t) > d_B^a(q_t) = d^a(q_t)$ as $q^N < q^{Mb} < 1 - \hat{q}^b(\gamma)$. Thus $d^b(q_t) > d^a(q_t)$ and $q_{t+1} < q_t$.

From this, we conclude that the cultural dynamics converges towards q^{Mb} .

b) The second situation is when $\hat{\gamma} < \gamma < \gamma_2$. In such a case, $q^{Mb} < \hat{q}^a(\gamma) < 1 - \hat{q}^b(\gamma)$. Then we have

- for $q_t < q^{Mb}$, $d^b(q_t) = d_B^b(q_t) < d_N^a(q_t) = d^a(q_t)$; thus $q_{t+1} > q_t$.
- for $q^{Mb} \leq q_t \leq \hat{q}^a(\gamma)$, $d^b(q_t) = d_B^b(q_t) > d_N^a(q_t) = d^a(q_t)$; thus $q_{t+1} < q_t$.
- for $\hat{q}^a(\gamma) \leq q_t \leq 1 - \hat{q}^b(\gamma)$, $d^b(q_t) = d_B^b(q_t) > d_N^a(q_t) > d_B^a(q_t) = d^a(q_t)$; thus $q_{t+1} < q_t$.
- for $1 - \hat{q}^b(\gamma) \leq q_t$, $d^b(q_t) = d_B^b(q_t) > d_N^a(q_t) > d_B^a(q_t) = d^a(q_t)$ (the second inequality comes from the fact that $q^N < q^{Mb} < \hat{q}^a(\gamma) < 1 - \hat{q}^b(\gamma)$). Thus $q_{t+1} < q_t$.

From this we conclude again that the cultural dynamics converge towards q^{Mb} . Finally in such a regime we have trait b associated to a "broad" identity and trait a associated to a "narrow" identity.

iv) case $\gamma_2 < \gamma < \gamma_3$. Consider then $\gamma'_2 \in (\gamma_2, \gamma_3)$ such that $\hat{q}^a(\gamma) = q^N$

a) A first situation is when $\gamma_2 < \gamma \leq \gamma'_2$. Then we necessarily have $q^N < \hat{q}^a(\gamma) < q^{Mb} < 1 - \hat{q}^b(\gamma)$. It follows that:

- for $q_t \leq q^N < \hat{q}^a(\gamma)$, $d^b(q_t) = d_B^b(q_t) < d_N^b(q_t) < d_N^a(q_t) = d^a(q_t)$, thus $q_{t+1} > q_t$.
- for $q^N < q_t \leq \hat{q}^a(\gamma)$, $d^b(q_t) = d_B^b(q_t) < d_N^a(q_t) = d^a(q_t)$ as $\hat{q}^a(\gamma) < q^{Mb}$, thus $q_{t+1} > q_t$.
- for $\hat{q}^a(\gamma) < q_t \leq q^{Mb}$, $d^b(q_t) = d_B^b(q_t) > d_B^a(q_t) = d^a(q_t)$ as $q^B < q^N < \hat{q}^a(\gamma)$, thus $q_{t+1} < q_t$.
- for $q^{Mb} < q_t \leq 1 - \hat{q}^b(\gamma)$, $d^b(q_t) = d_B^b(q_t) > d_B^a(q_t) = d^a(q_t)$ as $q^B < q^N < \hat{q}^a(\gamma) < q^{Mb}$, thus $q_{t+1} < q_t$.
- for $1 - \hat{q}^b(\gamma) < q_t$, $d^b(q_t) = d_B^b(q_t) > d_N^a(q_t) > d_B^a(q_t) = d^a(q_t)$ as $q^N < \hat{q}^a(\gamma) < q^{Mb} < 1 - \hat{q}^b(\gamma)$, thus $q_{t+1} < q_t$.

From this we conclude that the cultural dynamics should converge towards $\hat{q}^a(\gamma)$. and finally in such a regime we have trait b associated to a "broad" identity and trait a associated to a "narrow" identity or a "broad" identity.

b) The second situation is when $\gamma'_2 < \gamma \leq \gamma_3$. In such a case we have $q^B < \hat{q}^a(\gamma) < q^N < q^{Mb} < 1 - \hat{q}^b(\gamma)$

- for $q_t \leq q^B$, $d^b(q_t) = d_B^b(q_t) < d_B^a(q_t) < d_N^a(q_t) = d^a(q_t)$, thus $q_{t+1} > q_t$.
- for $q^B < q_t \leq \hat{q}^a(\gamma)$, $d^b(q_t) = d_B^b(q_t) < d_N^a(q_t) = d^a(q_t)$, (as $q_t \leq \hat{q}^a(\gamma) < q^{Mb}$) thus $q_{t+1} > q_t$.
- for $\hat{q}^a(\gamma) < q_t \leq q^N$, $d^b(q_t) = d_B^b(q_t) > d_B^a(q_t) = d^a(q_t)$ (as $q^B < \hat{q}^a(\gamma) < q_t$), thus $q_{t+1} < q_t$.
- for $q^N < q_t \leq 1 - \hat{q}^b(\gamma)$, $d^b(q_t) = d_B^b(q_t) > d_B^a(q_t) = d^a(q_t)$ (as $q^B < q^N < q_t$), thus $q_{t+1} < q_t$.
- for $1 - \hat{q}^b(\gamma) < q_t$, $d^b(q_t) = d_B^b(q_t) > d_N^a(q_t) > d_B^a(q_t) = d^a(q_t)$ as $q^N < 1 - \hat{q}^b(\gamma) < q_t$, thus $q_{t+1} < q_t$.

From this we conclude again that the cultural dynamics converge towards $\hat{q}^a(\gamma)$. In such a regime, trait b is associated to a "broad" identity and trait a is indifferent between a "narrow" identity or a "broad" identity.

v) case $\gamma \geq \gamma_3$. In such a case we have $\hat{q}^a(\gamma) \leq q^B < q^N < q^{Mb} < 1 - \hat{q}^b(\gamma)$

- for $q_t \leq \hat{q}^a(\gamma)$, $d^b(q_t) = d_B^b(q_t) < d_B^a(q_t) < d_N^a(q_t) = d^a(q_t)$, thus $q_{t+1} > q_t$.
- for $\hat{q}^a(\gamma) < q_t \leq q^B$, $d^b(q_t) = d_B^b(q_t) < d_B^a(q_t) = d^a(q_t)$, (as $q_t \leq \hat{q}^a(\gamma) < q^{Mb}$), thus $q_{t+1} > q_t$.
- for $q^B < q_t \leq q^N$, $d^b(q_t) = d_B^b(q_t) > d_B^a(q_t) = d^a(q_t)$ (as $\hat{q}^a(\gamma) \leq q^B < q_t$), thus $q_{t+1} < q_t$.
- for $q^N < q_t \leq 1 - \hat{q}^b(\gamma)$, $d^b(q_t) = d_B^b(q_t) > d_B^a(q_t) = d^a(q_t)$ (as $q^B < q^N < q_t$), thus $q_{t+1} < q_t$.
- for $1 - \hat{q}^b(\gamma) < q_t$, $d^b(q_t) = d_N^b(q_t) > d_N^a(q_t) > d_B^a(q_t) = d^a(q_t)$ (as $q^N < 1 - \hat{q}^b(\gamma) < q_t$), thus $q_{t+1} < q_t$.

From this we conclude that the cultural dynamics converge towards q^B . In such a regime, we have a "Full Broad Identity" structure with both cultural traits associated to the broad identity $B = \{a, b\}$.

Collecting the different results provides figure (6). **QED**