Identity and the Dynamics of Preferences

Davide Dragone* Michele Viviani†

FIRST VERSION 13/12/2006
THIS VERSION 21/5/2007

Abstract

Some recent economic contributions have studied individual identity in terms of explicit choices and social categories to which a person belongs to. According to Social Psychology, identity is also the result of a process influenced by self-regulation mechanisms. We model endogenous identity-dependent preferences as the dynamic result of two mechanisms: environmental pressure and the persistent effect of past socialization in the adaptation to new environments

We apply this model to environments where the agent must trade-off conflicting utility functions, such as material and non-material payoffs, or self-interest and other-regarding preferences. The model shows that heterogeneity in individual preferences, besides being the result of socialization, cultural transmission and environmental incentives, critically hinges on the feed-back of behaviour on preferences[1].

Keywords: Identity, Endogenous Preferences, Self-Regulation, Intrapersonal Conflict, Dual-Self, Multi-Cultural, Optimal Control.

JEL classification: A12, C61, D91, Z13

*Department of Economics, University of Bologna, Strada Maggiore 45, 40125, Bologna, Italy; e-mail: davide.dragone@unibo.it

†Department of Economics, University of Bologna, Piazza Scaravilli 2, 40125, Bologna, Italy; e-mail: michele.viviani2@unibo.it

[1]We thank Giacomo Calzolari, Emanuela Carbonara, Matteo Cervellati, Tiziana Mancini, Arsen Palestini, Federico Perali, Pier Luigi Sacco, Claudio Zoli for useful suggestions.

τὸ αντίξουν συμφέρον καὶ εκ τῶν διαφερόντον καλλίστην αρμονίαν καὶ πάντα κατ' έριν γίνεσθαι (Heraclitus, DK, B8)[2]

1 Introduction

Identity has recently received increasing attention in the economic literature. In particular it has been used to highlight the effect of social processes, like the belonging to social categories, on individual behavior. In this paper, we propose a model of endogenous identity-dependent preferences to foster two strictly related objectives.

The first one is to link three recent streams of the economic literature: the literature on identity-dependent preferences (Akerlof, Kranton, 2000), the literature on endogenous preferences (Becker, 1996) and the literature on the dependence of individual preferences on socialization in a multicultural environment (Bisin et al., 2006).

The second goal is to propose a compact analytical representation of some critical perspectives and results of the social-psychological theories, with special interest to the environments where an individual is subject to different (and potentially conflicting) identity references. This approach allows us to derive a theoretical explanation of the emergence of identity-related pathological behaviours (like extremism) that is consistent with the current Social Psychology literature.

In this paper we introduce a dynamic perspective where identity results endogenously from behaviour and previous experiences. To do so, we introduce feed-back mechanisms that drive the endogenous change of identity and, consequently, the identity-dependent preferences of the decision-maker. More specifically, we consider two distinct dynamic mechanisms that operate on the self-regulation of individual identity: environmental-pressure and anchoring.

The first mechanism refers to the fact that individual identity changes over time according to a pressure to conform to the requests of the environment where the individual lives. The second mechanism, anchoring, is introduced in order to consider the case of an individual that has developed a certain identity within a given social environment that, for some reason, changes. This can be the case of a second generation immigrant that has been socialized by its ethnic group and that, at a certain point of his life, begins to interact with different social groups. Or it can be the case of a

[2] *What opposes unites, and the finest attunement stems from things bearing in opposite directions, and all things come about by strife.*

worker that has been educated to a specific organizational culture and, at a certain point of its career, changes job and goes to a work place that has significantly different organizational values, norms or procedures (Mancini, 2001, 2006). As the literature reports, in all these cases the previous experiences can "anchor" the process of adaptation to the new environment and thus they affect the process of identity formation (Erikson, 1982, Greenwald, 1980).

The paper is structured as follows. In next Section we review the literature on identity and endogenous preferences. In Section III we study the evolution of identity as a deterministic optimal control problem, abstracting from uncertainty, bounded rationality and from strategic interaction. This framework is helpful to explain what factors influence the behavior of people over time and it suggests that heterogeneity in individual identity (and preferences), besides being the result of socialization and cultural transmission, explicit choices and environment incentives, is critically influenced by the individual history of people. In Section IV we provide a dual-self application in which we focus on the intrapersonal conflict that arises when people are simultaneously subject to different social norms. This allows to consider conflicts occurring at two different levels. The first one refers to a conflict *between social environments* characterized by different (exogenous) identity references The second one, specifically expressed by the dual-self assumption, introduces also the conflict *within a given environment* as expressed by the weight given to different prescriptions. This setting provides an intuitive interpretation of the relation between identity and social prescriptions. It allows to introduce a classification of identity and behavior, showing why, in a dynamic perspective, an agent can behave in pathological or extreme ways. Section V concludes.

Our perspective develops the contribution by Akerlof and Kranton on identity by considering a process in which identity changes and, consequently, also the identity-dependent preferences of the agent change over time. The endogeneity we introduce is driven by self-regulation mechanisms and is quite different from Becker (1996), where the endogenous modification of the utility function is driven by the formation and consumption of a capital good that depreciates without explicit reference to conformist pressure or anchoring dynamics. By considering identity as a complex construct that is both the result of a conscious choice, as well as the unconscious result of the experiences of the past and the pressures of the environment, we believe that this paper can provide a complementary perspective to the economic contributions that consider identity as an explicit object of choice in a multi-cultural

environment, as in Bisin et al. (2006). To do so, we explicitly refer to three streams of literature in Social Psychology that consider identity as a cognitive process ruled by self-regulation mechanism of personal representation (preferences) and behaviour; as a social process made of (multiple and subjective) belonging to social groups and social roles (collective identity) and as a motivational process whose objective is to stabilize personal representation.

2 Literature

2.1 Identity in Economics

The economic literature deals with identity in several contributions[3].

Akerlof and Kranton (2000) and Bisin and et al. (2006) consider identity as the membership to social groups whose norms, costumes and culture are reflected in the individual preferences of the agents. In these contributions, identity is at the same time a motivation for individual behaviour and a choice of the individual. In Akerlof and Kranton (2000), identity is defined as a person's sense of self that depends on the belonging to specific social categories that determine specific behavioral prescriptions, a position that is consistent with the sociological contributions by Tajfel and Turner (1979). In this perspective identity affects the individual preferences and it is represented as an argument of the individual utility function. In their proposal identity can be chosen by the individual, it operates as an externality in people's interactions and it can be socially manipulated, for example through advertisement. The contributions by Bisin and his coauthors (Bisin, Verdier, 2001; Bisin, Topa, Verdier, 2004; Bisin, Patacchini, Verdier, Zenou, 2006) consider the relationship between selection of preferences within a population by focusing on the concepts of cultural transmission (socialization) and identity in multicultural environments. In Bisin, Patacchini, Verdier, Zenou (2006), the authors explicitly refer to individual identity in order to add to peer and parental pressure also the role of individual choice in ethnic identity determination. Specifically, individual identity is expressed by the weight attributed to the prescriptions of the minority culture Vs. the majority one. In a recent contribution, Bénabou and Tirole (2007) consider identity as a result of an individual investment in beliefs when they are uncertain about their preferences. This endogenizes the identity-related payoffs in a setting with incomplete information.

[3]For surveys, see Davies (2004) and Hill (2006).

2.2 Identity in Social Psychology

In Social Psychology, identity is not just the result of a choice, as in most of the economic contributions, but it can also be defined as the *process* through which people define their Self. Mancini (2001) proposes to organize the literature on identity into three streams: cognitive, social and motivational approaches.

Cognitive perspective. The cognitive perspective focuses on the process through which self-knowledge or self-representation is elaborated and organized at a cognitive level. From this perspective, the Self is considered to be an organized structure of self-representations that regulates the interaction between the person and the environment. These images are both the input and the output of a psychological process of self-regulation of behaviour (Markus, Wurf, 1987). This means that the individual builds her Self over time through "a continual process of moving towards and away from various kind of mental goal representation, and this movement occurs by a process of feedback control" (Carver, Scheir, 1998:2)[4].

Social perspective. The concept of identity we use in this paper refers to the concept of collective identity as proposed by Ashmore, Deaux and McLaughlin-Volpe (2004):

> "A collective identity is one that is shared with a group of others who have (or are believed to have) some characteristic(s) in common [...]. Such commonality may be based on ascribed characteristics, such as ethnicity or gender, or on achieved states, such as occupation or political party [...]. This shared position does not require direct contact or interchange with all others who share category membership; rather, the positioning is psychological in nature [...] defined here in terms of a subjective claim or acceptance by the person whose identity is at stake [...]. That is, although others may refer to one in terms of a particular social category, that category does not become a collective identity unless it is personally acknowledged as self-defining in some respect" [Ashmore et al., 2004:81].

Collective identity thus generalizes the more familiar concept of social identity (Tajfel, Turner, 1979) by referring to the subjective belonging to social groups and/or social roles of a person. The social perspective on

[4] A mental goal refers to any salient goal that motivates individual behaviour and that is determined by, for example, the preferences of the individual, by the social environment and the kind of task the individual has to carry over.

identity emphasizes the role of the social environment within which identity is developed. Within this perspective we can find both the theory proposed by Mead (1932) and the Social Identity Theory by Tajfel and Turner. Mead's contribution considers, in a nutshell, the self as a social construct that results from repeated interactions with other people belonging to the same group or community. Rosemberg (1988) and Turner (1968, 1987) developed on these premises, arguing that the social nature of the self depends on the different roles within which people identify themselves and the different environments they have lived in. Turner, in particular, underlines that the concept of the self is a selective organization of values and models of behaviour that are learned through social interaction. In this perspective the social norms (or the culture) of a community are the building blocks for the construction of self. The Social Identity Theory focuses on the elements of the self that depend on the belonging to particular groups or social categories that are determined, for example, on ethnic, gender, professional or religious bases.

Motivational perspective. The motivational approaches to identity consider the psychological and social needs as the engine for the construction of identity. A motivation for the management of identity is the search for coherence and continuity in the perception of self. The main contributions in this field are Erikson (1968, 1982) and Marcia (Marcia et al., 1993). Erikson considers the formation of identity as a psycho-social process that integrates the abilities, beliefs and childhood experiences with the individual expectations and the requests of the social environment. Marcia develops Erikson's proposal in the "Paradigm of Identity States". This perspective suggests the importance of a stable identity as a psychological need, an argument that is taken also by the cognitive perspective (Greenwald, 1980) when considering a tendency to "conservatorism" in the representation of individual identity. The motivational approach thus suggests that the evolution of identity and individual behaviour are driven, as the social perspectives claim, by a conformism pressure that depends on the environment the individual is embedded into. It also points out that there exists a need for stability in the identity constructs. In an multi-cultural environment or when the environment changes (which are situations that often occur in the present society), these two drives can be conflicting, a situation that is commonly considered by the psychological and sociological literature on contemporary multi-cultural contexts to be critical for the individual management of identity (Baumeister, Muraven, 1996; Berger, Berger and Keller, 1983, Bauman 2002).

2.3 Endogenous preferences

The dynamic aspect of individual choice and behaviour has been introduced in the economic literature by Becker (Becker, 1996) as the endogenous change of a parameter of the utility function. More specifically, Becker considers the formation and consumption of a capital good (consumption capital, social capital, human capital or a habit) that changes over time. As this affects the preferences of the agent, the dynamics of this capital good also affects the intertemporal behavior of the agent.

Other approaches on endogenous preferences do not focus on the individual: they consider the selection of a specific set of preference from a population perspective. Bisin, Topa, Verdier (2004) distinguish three approaches. The first one concerns direct evolutionary selection mechanisms of specific behavioral rules like, for example, altruistic versus. cooperative strategies within a population (Gintis, 2003). Other scholars, like Guth and Yaari (1992) or Ok and Vega-Redondo (2001), have proposed models of indirect evolutionary selection where, instead of behavioral strategies, the selection occurs on individual preferences. A third perspective focuses on cultural transmission mechanisms in which the diffusion of a specific pattern of preference distribution depends on direct socialization mechanisms (Bisin et al., 2001, 2004, 2006)

Contributions from the psychological and sociological literature observe that preferences may also change over time as a result of the management of intrapersonal conflicts. Interpersonal conflict means that an individual can evaluate an action according to different, potentially conflicting criteria (Elster, 1986, Schelling, 1978)[5]. Ainslie (1992), for example, suggests that the intertemporal behavior of people is driven by the interaction of sequential motivational states within the person and that this can be the cause of dynamically inconsistent behavior. This perspective has also been adopted in some economic contributions on intertemporal choice where the same set of alternatives is assumed to be differently ranked according to a long or a short run perspective (Bénabou, Pycia, 2002; Bénabou, Tirole, 2004, Fudenberg, Levine, 2005; Gul, Pesendorfer, 2001; Laibson, 1997; O'Donoghue, Rabin, 1999). Recently, this multiple-self approach to decision-making has been justified on the basis of the literature on the cognitive processes that occur within the brain. Accordingly, the intrapersonal conflict would not occur only between sequential selves over time, but also between affective or deliberative systems (Lowenstein, O'Donoghue, 2004) or automatic and

[5]See Freud (1924) for a seminal reference in psychology, Stark and Deutsch (2004) and Camerer et al. (2005) for some recent contributions in, respectively, sociology and cognitive sciences.

controlled processing of environmental stimuli (Benhabib, Bisin, 2004) that simultaneously interact in the brain and affect individual behavior.

3 The model

3.1 The static objective function

As in Akerlof and Kranton (2000), we consider identity-dependent preferences and we represent (collective) identity as a parameter affecting the instantaneous utility function of the agent. Under standard assumptions, we assume that the agent's objective function depends on the behavior of the agent, on her identity and on the environment in which she is living. This means that, given a compact set $X \in R^n$, the alternatives $x(t) \in X$ can be ranked according to the following utility function

$$v(x(t), a(t), a^e).$$

The utility function is continuously twice differentiable and it is a negative definite quadratic form; we assume that the standard assumptions for utility functions hold, except for the fact that satiation is admitted[6]. The term $a(t) \in [0, 1]$ represents the collective identity of the agent at time t and $a^e \in [0, 1]$ is a given environment-dependent identity reference. Through all the paper, we assume that neither time nor the agent affect the environmental reference.

3.2 The dynamics of identity: environmental pressure and anchoring

In the economics contributions on identity, identity is essentially a static choice and a static parameter in the individual preferences. Nevertheless, the Social Psychological literature argues that identity should also be considered as the result of a dynamic process through which the individual defines her psychological belonging to a social group. This implies that the individual behavior changes over time as a result of the interaction with the environment. The theoretical model we propose studies the optimal path of behavior and its effect on individual preferences, as represented by the variation of the identity parameter $a(t)$ over time. To introduce the dynamics of identity as a result of behaviour, we consider feed-back mechanisms that drive the endogenous change of identity and of the preferences of the

[6] See the dual-self application in next Section for details and interpretation.

decision-maker. As it is commonly assumed in Economics, behaviour depends on preferences; moreover we also allow preferences to depend on behavior, as they are endogenously influenced, via the change in the value of a, by past choices.

More specifically, we assume the existence of two dynamic mechanisms, that we call environmental-pressure and anchoring. The first mechanism refers to the fact that individual identity depends on a pressure to conform to the requests of the environment where the individual lives. The *environmental pressure* mechanism $f(a(t), x(t), a^e)$ is driven by a welfare-improving criterion that drives identity in order to adapt it to the social environment the agent is living in. Analytically, this mechanism is based on the partial derivative of the objective function with respect to the identity parameter $a(t)$, so that

$$f(x(t), a(t), a^e) = v_a(x(t), a(t), a^e). \tag{1}$$

If this derivative is positive (resp. negative), this means that, given the environment and the behavior of the agent, an increase (resp. decrease) in a would increase her well-being.

The second mechanism, *anchoring*, is thus useful to consider the case of an individual that has lived in a social environment that is significantly different from the one the agent is actually living in. Consistent with the motivational perspective on identity, we consider the dynamics of identity as a process directed at stabilizing the personal representations of the self. So we assume that the identity references of the previous environment never lose their motivational power and that they operate as an anchor for the collective identity representations of the individual.

Consider a second generation immigrant that has been socialized to the culture of its ethnic group (or family) and develops her identity with respect to that specific environment. Alternatively, consider the case of a person that has been working in a firm with a specific organizational culture. At a certain point in time, the agent moves to a different social environment: the second generation immigrant goes to work or study in a place where a different identity reference exists, the worker changes workplace, etc. In these cases, we assume that the previous socialization experiences create an identity anchor $a^m \in [0, 1]$ (the superscript m stands for memory) to which individual identity tends to come back according to the anchoring function $g(a(t), a^m)$. This function does not depend on individual behavior x, it is continuously differentiable and such that $g(a(t), a^m) \gtreqless 0$ when $a(t) \gtreqless a^m$.

The combined effect of the environmental pressure and the anchoring

mechanism determines the transition equation for identity as follows:

$$\dot{a}(t) = f(x(t), a(t), a^e) + g(a(t), a^m) \tag{2}$$

Note that the environmental reference a^e and the anchor reference a^m are conceptually different. The environmental reference has both a static and a dynamic effect: the static effect concerns the direct impact on utility of the prescribed identity reference, while the dynamic effect represents the conformist pressure that induces identity to change to adapt it to the environmental reference. The anchor reference, on the contrary, does not depend on the current environment and it does not enter the instantaneous utility function: its effect only operates on the dynamics of identity and we assume, for simplicity, that this effects never fades away.

3.3 Static solution

Before studying the dynamic solution, some remarks on the properties of the static solutions are useful for later comparison. To do so, we consider the optimal choice of x as if no self-regulation mechanism existed. In other words, we consider the static case where both the type of agent (i.e. her identity a) and the environmental reference a^e, are exogenously given. The static problem is the following:

$$\max_{x \in X} v(x, \bar{a}, a^e) \tag{3}$$

$$\text{s.t. } M = px \tag{4}$$

with $\bar{a} \in [0, 1]$, for any t, $p \in R^n$ is a given vector of prices and M a fixed available endowment. Given the strict concavity of the function, the internal solution $x^s = x^s(\bar{a}, a^e, p, M)$ to the static problem is unique.

For later comparison note that, could the individual choose her identity and the corresponding behavior, she would choose the combination of identity and behavior that allows reaching the global maximum (if it exists) of the objective function. This (static) welfare-maximizing combination of identity and behavior (a^G, x^G) depends on the environmental reference a^e and, if belonging to the compact set $[0, 1] \times X$, it satisfies the following conditions:

$$v_x(a^G, x^G, a^e) = 0 \tag{5a}$$

$$v_a(a^G, x^G, a^e) = 0 \tag{5b}$$

As we will see, there are conditions that allow this global solution of the static maximization problem to be also the steady state solution of the

dynamic problem. This means that, even though we do not allow the agent to directly choose her identity parameter a, there are conditions that allow the path of choices to correctly feed-back on preferences in order to ensure the highest utility to the agent.

3.4 Dynamic solution

By focusing on a dynamic framework, we want to study the evolution of identity due to the feed-back effects of choices on preferences. This means that there is a circular relationship between preferences and choices. As a consequence the optimal choice in the dynamic framework can be different than the corresponding optimal static choice, depending on the persistence of past experiences.

In order to focus our attention on the endogenous change of preferences over time, we consider a simplified framework in which there is no interaction with other agents; furthermore no uncertainty or ambiguity are allowed for, so that our results do not hinge on informational issues. In an infinite-time horizon, the optimal path of behavior must solve the following program:

$$\max_{\{x(t)\}} \int_{t=0}^{\infty} e^{-rt} v(x(t), a(t), a^e) dt \tag{6a}$$

$$\text{s.t. } \dot{a}(t) = f(x(t), a(t), a^e) + g(a(t), a^m) \tag{6b}$$

$$M = px(t) \tag{6c}$$

$$a(0) = a_0 \tag{6d}$$

$$a(t) \in [0, 1], \forall t \tag{6e}$$

$$x(t) \in X, \forall t \tag{6f}$$

where a^e and a^m are given, $f(x(t), a(t), a^e)$ represents the dynamic effect of the environmental pressure, $g(a(t), a^m)$ is the component based on anchoring and a_0 is a given initial identity. We also require the general transversality condition for infinite-time horizon problems, $\lim_{t\to\infty} e^{rt} H(t) = 0$, to be satisfied.

The current-value Hamiltonian of the dynamic problem is the following (the time index is omitted):

$$H = v(x, a, a^e) + \lambda[v_a(x, a, a^e) + g(a, a^m)] \tag{7}$$

where λ is the costate variable and $f(x(t), a(t), a^e)$ has been replaced accord-

ing to (1). The corresponding canonic equations are:

$$\dot{a} = v_a(x, a, a^e) + g(a, a^m) \tag{8}$$

$$\dot{\lambda} = [r - v_{aa}(x, a, a^e) - g_a(a, a^m)]\lambda - v_a(x, a, a^e) \tag{9}$$

For interior solutions x^d, the first order conditions on the Hamiltonian function are:

$$H_x = v_x(x, a, a^e) + \lambda v_{ax}(x, a, a^e) = 0. \tag{10}$$

Let $F(a, \lambda)$ be the function that, given a and λ, assigns the optimal behavior (or choice) x^d according to equation (10).

3.5 Properties of the steady states

Given the general form of the problem, we cannot define the number and positions of the steady states without a specific objective function and a specific dynamics of identity. Nevertheless we are able to point out three properties of a steady state pair (a^*, x^*). The first one concerns the role of the environmental pressure and the anchoring mechanism in identifying a steady state solution. The second one points out the instability of the steady state solutions. The third remark identifies a class of steady state solutions that coincide with the global optimum of the static objective function and it points out that the persistence of anchoring leads to second best outcomes.

First remark. An internal steady state (x^*, a^*) must satisfy equation (8) and (9) with equality. The first one,

$$f(x^*, a^*, a^e) + g(a^*, a^m) = 0, \tag{11}$$

implies that the environmental and the anchoring mechanisms must compensate each with the other, a condition that can be interpreted saying that a steady state requires the environmental pressure on identity to have different sign (but equal strength) with respect to the memory effect of anchoring.

A *second remark* concerns the stability properties of the steady states, properties that can assessed by eliminating the control variable x and expressing the dynamic system as a function of the state and costate variables. After some computations (see the appendix the details), the dynamic system becomes:

$$\dot{a} = v_a(F(a, \lambda), a, a^e) + g(a, a^m) \tag{12a}$$

$$\dot{\lambda} = (r - v_{aa}(F(a, \lambda), a, a^e) - g_a(a, a^m))\lambda - v_a(F(a, \lambda), a, a^e). \tag{12b}$$

We can now determine the local stability properties of the steady states that satisfy system (12) with equality by considering the following Jacobian matrix:

$$
\begin{aligned}
J &= \begin{bmatrix} v_{aa} + v_{ax}F_a + g_a & v_{ax}F_\lambda \\ -g_{aa}\lambda - v_{aa} - v_{ax}F_a & r - g_a - v_{aa} - v_{ax}F_\lambda \end{bmatrix} \\
&= \begin{bmatrix} v_{aa} - \frac{(v_{ax})^2}{H_{xx}} + g_a & -\frac{(v_{ax})^2}{H_{xx}} \\ \frac{(v_{ax})^2}{H_{xx}} - v_{aa} - g_{aa}\lambda & r - g_a - v_{aa} + \frac{(v_{ax})^2}{H_{xx}} \end{bmatrix}
\end{aligned} \tag{13}
$$

Since $Tr(J) = r > 0$, at least one eigenvalue has a positive real part, indicating that the steady states of the dynamic problem are unstables (i.e. saddles or sources), but that in no case it is possible to get a sink (that would be associated with two eigenvalues with negative real parts) or a cycle (purely imaginary eigenvalues).

The *third remark* focuses on the utility that can be reached in the steady state, showing that suboptimal results occur because of the long-lasting effect of the anchoring mechanism, and not because of any form of dynamic inconsistency. In fact, by considering a standard intertemporal utility with constant discount rate $r \in [0, 1]$, we are explicitly ruling out the possibility that dynamic inconsistent behavior (i.e. divergence between planning and implementation of the plan) arises. To see it, consider a feasible steady state (a^*, x^*) that simultaneously satisfies with equality the first order condition (10) and the canonical equations (8) and (9). Consider the special case in which the anchoring effect is nil, $g(a^*, a^m) = 0$, a condition that occurs when the steady state identity and the anchoring references are the same, $a^* = a^m$, or when there is no persistence of previous socializations. Since (a^*, x^*) is internal, it satisfies the following system of equations:

$$v_x(a^*, x^*, a^e) = 0 \tag{14a}$$

$$v_a(a^*, x^*, a^e) = 0. \tag{14b}$$

meaning that the steady state pair (a^*, x^*) coincides with the global maximum of the static objective function (a^G, x^G). In other words, when anchoring has no effect, the optimal path of choices and identities is such that the optimal identity a^G and the corresponding optimal behavior x^G (given the environmental reference a^e) is reached. By reversing the argument, this means that, whenever anchoring is not nil in the steady state, we should observe a suboptimal steady-state solution. This is consistent with the idea that the evolution of identity is a process that integrates the experiences of life, with specific reference to the environmental references a person has been exposed

to (Erikson, 1968, 1994). When past references persist, the adaptation of the agent to the new environment is not optimal. An implication is that the evolution of identity is a truly environment-dependent process since people, even if subject to identical environmental incentives, will develop different identities as a consequence of their personal history and, specifically, to the persistence of the previous environmental references.

In order to get a deeper understanding of the factors that influence the optimal path of choices and the endogenous evolution of identity, in next section we propose an application that shows how certain steady states, beside being generically suboptimal, can be labelled as pathological solutions in which the joint effect of reinforcement and anchoring induce to people "overshoot" and (rationally) make extreme choices.

4 A Dual-Self application

In this section we present an application to the model presented in the previous section. This allows for an intuitive interpretation and classification of collective identity in an environment where multiple social norms (or cultural prescriptions) coexist. The application is also useful to introduce a classification of behavior and to explain the conditions that induce the emergence of pathological or extreme behaviors.

Here we consider a social norm as a ranking over the possible behaviors (i.e. it prescribes how to evaluate a behavior[7]). Under the usual assumptions, a social norm can be represented as a utility function; the only remarkable difference with the usual utility functions is that satiation is admitted when behavior perfectly satisfies the norm. Typically, in a multicultural society people are simultaneously subject to different social norms or prescriptions. This can be the case of a second generation immigrant that is both subject to religious and secular prescriptions, or a worker that is both required to cooperate and to compete with her colleagues.

In this context, we interpret identity a as the weight the individual gives to each social norm she is subject to and we want to understand, *i)* what the behavior of the agent is in a static framework, *ii)* how identity changes over time as a consequence of reinforcement and *iii)* how behavior changes over time.

[7]For an alternative approach based on descriptive norms, instead of prescriptive ones, see Bicchieri (2006).

To represent the prescriptions associated to each social norm, we consider two exogenous, rational and different preference orderings, represented by the (sub)utility functions $n(x)$ and $m(x)$. These functions are assumed to be continuously differentiable and strictly concave. As there exist an optimal behavior according to each set of prescriptions, the sub-utility functions admit satiation when the social norm is perfectly satisfied, which occurs in, respectively, x^n and x^m.

The existence of different social norms justifies the existence of different rankings on the same set of choice, which can be interpreted in terms of intrapersonal conflict. We assume that the decision-maker evaluates the available alternatives $x \in X$ by considering the following linear combination of two sub-utility functions

$$an(x) + (1-a)m(x), \tag{15}$$

with $a \in [0,1]$ that represents how the agent trades-off the different social norms. We interpret a as the collective identity of the decision-maker, as it represents the relative importance of each social norm in the overall evaluation of outcomes.

Given that a is defined over a continuum of values, we can easily classify the identity and the behavior of the decision-maker. When $a = 0$ and $a = 1$, the agent is said to have a *polar identity*, meaning that she is subject to only one set of social norms. Analogously, we can define as *polar choices* those that correspond to x^n and x^m, i.e. the optimal behavior prescribed according to each subutility function. If we consider a simplified problem where $X = R$, we can also introduce a classification on all choices. More specifically, we define a *compromise choice* as a choice x that lies between the optimal action of each social norm (i.e. x^m and x^n), meaning that the individual is mediating between the two norms. As we will see, in a static framework a polar identity is always associated with a polar behavior; nevertheless in the dynamic model this property in general does not hold. In particular, we show that a new class of behaviors those that stay out of the interval $[x^m, x^n]$ -we call them *extreme-choices*- can emerge even if identities are not polar[8].

We assume also that there exists a stable, exogenous identity reference value a^e given by the (social) environment and that deviating from this reference yields disutility to the agent. This disutility can be due to the actions

[8]Without loss of generality, we can assume $x^m < x^n$.More generallly, we define a compromise choice a value of x such that $|x - x^u| + |x - x^m| = |x^u - x^m|$ (with $x \neq x^u$ and $x \neq x^m$), a choice is polar if the previous equality holds and $x = x^u$ or $x = x^m$. A choice is extreme if x is such that $|x - x^u| + |x - x^m| > |x^u - x^m|$.

of people that react to the deviant identity of the agent (see Akerlof, Kranton, 2000 for the case of gender identity) or peer-pressure, but it could also be interpreted as the consequence of an internal (psychological) sanctioning mechanism, such as guilt. We assume this direct disutility cost to depend only on the distance between a and a^e and we represent it with the convex cost function $c(a, a^e)$.

The objective function of the agent is given by the aggregation of the dual utility function with the disutility cost of deviating from the given exogenous reference a^e as follows[9]:

$$v(a, x, a^e) = an(x) + (1 - a)m(x) - c(a, a^e) \tag{16}$$

This implies that the environmental pressure function in the transition equation is given by:

$$f(x, a, a^e) = [n(x) - m(x)] - c_a(x, a^e). \tag{17}$$

The first part of the expression (17) is the difference in the relative desirability that the two preference orderings induce. The second one represents the conformist effect and is given by the marginal cost of deviance from the environmental reference a^e. This implies that the disutility cost of deviance does not only have a direct impact on the agent's well being, but it also has an endogenous effect on the evolution of identity.

As anticipated in the previous section, we want to study also the effects of a change in the social environment. To do so, we assume that there exists an exogenous reference anchor a^m that influences on the evolution of identity through the function $g(a, a^m)$. This implies that the dual-self application allows to consider conflicts occurring at two different levels. The first one refers to a conflict *between social environments* characterized by different (exogenous) identity references The second one, specifically expressed by the dual-self assumption, introduces also the conflict *within a given environment* as expressed by the weight given to different prescriptions.

[9]Note that this function is strictly concave in both arguments, as $v_{xx} = au_{xx} + (1 - a)m_{xx} < 0$ and $v_{aa} = -c_{aa} < 0$, but that this does not guarantee the joint concavity in both variables, that is to be explicitly assumed. When this is not the case, boundary solutions -whose position critically depends on the size of the compact set $[0, 1] \times X$- emerge.

4.1 The static problem

In a static environment in which $a(t) = \bar{a}$ for any t, the agent solves the following problem:

$$\max_{x \in R} v(x, \bar{a}, a^e) = \bar{a}n(x) + (1 - \bar{a})m(x) - c(\bar{a}, a^e) \tag{18}$$

some remarks on the static solution $x^s = x^s(\bar{a}, a^e)$.

1. x^s depends, as one would expect, only on the partial derivative of the objective function

$$v_x = \bar{a}n_x + (1 - \bar{a})m_x = 0. \tag{19}$$

   Note that the disutility term has no effect on the optimal choice so that, in the static framework, having an identity different from the environmental reference a^e simply determines a fixed unavoidable cost of deviance.

2. Since the foc can be written as $\frac{n_x(\bar{a},x^s)}{m_x(\bar{a},x^s)} = -\frac{1-\bar{a}}{\bar{a}}$, the term $\frac{n_x}{m_x}$ computed in x^s is negative for any $\bar{a} \in (0, 1)$[10], indicating that x^s must stay in a position where the marginal sub-utility functions have different signs. This means that the optimal choice must trade-off the marginal increases of one sub-utility function with the marginal decrements of the other sub-utility function. We interpret a solution with such a property as *a compromise solution* because it stays between the two optimal points: $x^s \in (x^m, x^n)$

3. When the individual has a polar identity (i.e. $\bar{a} = 1$ or $\bar{a} \in 0$), then also the optimal static solution is polar: $x^s = x^n$ (or $x^s = x^m$)

4.2 The dynamic problem

The current-value Hamiltonian of the dynamic problem is the following (the time index is omitted when no confusion arises):

$$H = an(x)+(1-a)m(x)-c(a, a^e)+\lambda[n(x)-m(x)-c_a(a, a^e)+g(a, a^m)] \tag{20}$$

[10]The expression $\frac{u_x}{m_x}$ can be interpreted in terms of marginal rate of substitution. This can be seen considering, for example, an objective function such as $v(\bar{a}, x, y, a^e) = \bar{a}u(x) + (1 - \bar{a})m(y) - c(\bar{a}, a^e)$ and a budget constraint like $p_x x + p_y y = w$, where x and y are consumption goods, p_x and p_y their prices and w the individual endowement.

where λ is the costate variable and, according to the discussion in the previous paragraph, $f(x, a^e)$ has been substituted by $n(x) - m(x) - c_a(a, a^e)$. We also require the general transversality condition for infinite-time horizon problems, $\lim_{t\to\infty} e^{rt}H(t) = 0$, to be satisfied.

For interior solutions x^d, the first order conditions on the Hamiltonian function are:

$$H_x = (a + \lambda)n_x + (1 - a - \lambda)m_x = 0. \tag{21}$$

The corresponding canonic equations are:

$$\dot{a} = n(x) - m(x) - c_a(a, a^e) + g(a, a^m) \tag{22}$$

$$\dot{\lambda} = (r + c_{aa}(a, a^e) - g_a(a, a^m))\lambda + m(x) - n(x) + c_a(a, a^e) \tag{23}$$

With respect to the static solution x^s, some remarks on condition (21) are noteworthy.

1. When $\lambda \neq 0$, the static and dynamic solutions are different.

2. Given a, the distance between the static and the dynamic optimal behavior depends on the value of the costate variable λ. Equation (23) tells us that such a value of λ depends on $m(x)-n(x)$. This implies that x^d depends on the objective function according to two channels. The first one is represented by the marginal sub-utility functions (as in the static case). The second one, which emerges only in a dynamic setting, is given by the relative desirability of behavior according to the two social norms. As a consequence, if an exogenous shock S occurred to one set of norms implying an increase in the relative desirability of this norm (e.g. if the new subutility function $\tilde{n}(x)$ were $\tilde{n}(x) = n(x) + S$), this shock would not have any influence in the static framework in which the agent cannot change her identity. Nevertheless, in a dynamic setting the shock would affect the optimal behavior of the agent via the effect on the costate variable.

3. If $\lambda < -a$ or $\lambda > 1 - a$, then the marginal rate of substitution in x^d is positive. This means that the dynamic solution stays in a zone where there is no conflict between the two sub-utility functions. For example, if $\lambda < -a$, it can be the case that the static utility would increase if x^d increased[11]. In a static framework this would induce x to increase. Yet, when we introduce self-regulation mechanisms, it can be optimal not to do so, as an increase in x^d could induce a change in a that negatively affects (given the value of λ) the dynamic objective function.

[11] This occurs when both sub-utility functions are increasing and, more specifically, when the following conditions hold: $u_x(a, x^d) > m_x(a, x^d) > 0$. See the Appendix for details.

4. A related consequence of the previous remark is that the optimal dynamic choice of x^d can lie outside the interval $[x^m, x^n]$. In other words, the agent can find it optimal to make *extreme* (dysfunctional) choices that do not mediate between the marginal effects of the two sub-utility functions.

4.3 Comments on the dual-self application

This dual application is useful for understanding the relation between identity and behavior over time by considering an individual that moves from a multicultural environment to another one. In a static framework, if the agent has a polar identity, the optimal behavior coincides with one of the two optima. If the agent has an intermediate identity, the optimal behavior will always lay between the optima of the two social norms. Nevertheless, in a dynamic framework we can also find optimal behaviors that do not mediate and do not coincide with the optimal prescriptions of each norm. This can be the case of religious extremism of a second generation immigrant that neither follows the religious prescriptions, nor the secular culture, and that behaves in such a way that she is exacerbating the prescriptions of both social norms. Another example concerns food consumption in which a person is typically simultaneously subject to the standards proposed by the family and those advertised by the mass-media. In this situation, pathological behavior means that the person finds it optimal too eat too little, or too much, with respect to both standards, as it is the case of anorexia and bulimia.

Our explanation for this behavior does not rely on self-control problems (Ainslie, 1991, Laibson, 1997). Indeed, in our model, what is planned will be faithfully implemented in order to reach the steady state solution which, as shown above, can not coincide with the combination of identity and behavior that best fits the current environment the agent lives in. According to the model we propose, this behavior is the result of the management of the simultaneous conflict between norms and the dynamic conflict between different social identity references. Since we consider full information, our explanation of pathological behavior is also different from the explanation based on self-signalling, which concludes that dysfunctional behavior can be observed when people are uncertain about their own identity and infer them from their past choices in order to signal to future selves the belonging to a certain type (Bénabou, Tirole, 2007).

5 Conclusions

This paper studies the dynamics of individual identity and the endogenous evolution of preferences by explicitly including some relevant insights from the Social Psychology literature. The goal is to improve the economic understanding of the processes that dynamically influence the behaviour of people over time by studying the evolution of identity-dependent preferences.

In the model we propose, identity is not an explicit choice of the agent, neither a static category she is ascribed to. We study identity as the result of a dynamic process that depends on self-regulation mechanisms. We propose and study two specific kinds of mechanisms. According to environmental pressure mechanism, identity changes over time in order to increase the well-being of the agent, given her past actions. The second mechanism, anchoring, refers to the persistence of old identity references and their effect on an individual living in an environment where a different reference exists.

As it turns out, identity is path-dependent and heterogeneity between agents is shown to critically hinge on environmental changes, as well as individual behavior. We identify conditions for internal steady states to correspond to global maxima, as well as conditions that determine paths of behaviour leading to polar identities. The model also shows that, when the individual anticipates the effects of current choices on identity and future behaviour, pathological behaviours can emerge, i.e. behaviours that are far from all the relevant social prescriptions. This result is due to the endogenous effect of actions on preferences, so that the agent chooses extreme behaviors in order to prevent her identity from changing.

The model we propose offers a complementary perspective for studying endogenous preferences, with respect to Becker's (1996) habit formation model, by explicitly introducing insights from Social Psychology. The dual-self application shows that the model can be usefully applied to a variety of decision-making contexts in which intrapersonal conflict occurs, as it is the case of the literature studying how people trade off material vs. non material gains, self-interest versus other regarding preferences, short-run vs. long-run concerns.

6 References

References

[1] Ashmore R.D., Deaux K., McLaughlin-Volpe T. (2004). An Organizing Framework for Collective Identity: Articulation and Significance of Multidimensionality, *Psychological Bulletin*, 130(1), 80-114.

[2] Ainslie G. (1992). *Picoeconomics: The Strategic Interaction of Successive Motivational States within the Person*, Cambridge: Cambridge University Press.

[3] Akerlof G.A., Kranton R.E. (2000). Economics and Identity, *Quarterly Journal of Economics,* 115(3), 715-753.

[4] Bauman Z. (2000). *Liquid Modernity*, Polity Press.

[5] Baumeister R.F., Muraven M. (1996). Identity as Adaptation to Social, Cultural and Historical Context, *Journal of Adolescence,* 19, 405-416.

[6] Becker G.S. (1996). *Accounting for Tastes*, Cambridge, Massachusetts, Harvard University Press.

[7] Bénabou R., Pycia M. (2002). Dynamic Inconsistency and Self-Control: a Planner–Doer Interpretation, *Economics Letters,* 77, 419–424

[8] Bénabou R., Tirole J. (2004). Willpower and Personal Rules, *Journal of Political Economy,* 112, 848-887.

[9] Bénabou R., Tirole J. (2007). *Identity, Dignity and Taboos: Beliefs as Assets*, IZA discussion paper 2583.

[10] Benhabib J., Bisin A. (2005). Modelling Internal Commitment Mechanisms and Self-Control: A Neuroeconomics Approach to Consumption-Saving Decisions, *Games and Economic Behaviour,* 52, 460-492.

[11] Berger P.L., Berger B., Keller H. (1983) *La Pluralizzazione dei Mondi di Vita*, in Sciolla L. (eds.): Identità, Percorsi di Analisi Sociologica, Rosemberg e Sellier, Torino.

[12] Bisin A., Patacchini E., Verdier T., Zenou Y. (2006). *Bend it Like Beckham, Identity, Socialization, Assimilation*, CEPR Discussion Paper No. 5662.

[13] Bisin A., Topa G., Verdier T. (2004). Cooperation as a Cultural Transmitted Trait, *Rationality and Society,* 16(4), 477-507.

[14] Bisin A., Verdier T. (2001). The Economics of Cutural Transmission and the Dynamics of Preferences, *Journal of Economic Theory*, 97, 298-319.

[15] Camerer C., Lowenstein G., Prelec D. (2005). Neuroeconomics: How Neuroscience can Inform Economics, *Journal of Economic Literature*, 43: 1, 9-64.

[16] Carver C.S., Scheier M. F. (1998). *On the Self-Regulation of Behaviour*, Cambridge: Cambridge University Press.

[17] Cavalli-Sforza L.L., Feldman M. (1981). *Cultural Transmission and Evolution: A Quantitative Approach*, Princeton, NJ: Princeton University Press.

[18] Davis J. B. (2006). Akerlof and Kranton on Identity in Economics: Inverting The Analysis, *Cambridge Journal of Economics*.

[19] Elster J. (1986). *The Multiple Self*. Cambridge: Cambridge University Press.

[20] Erikson E. H. (1968). *Identity Youth and Crisis*, New York: Norton & Co.

[21] Erikson E.H. (1982). *The Life Cycle Completed a Review*, New York: Rikan Enterprise.

[22] Freud S. (1924). *The Ego and the Id*, New York: W. Norton, 1962.

[23] Fudenberg D., Levine D. (2005). *A Dual Self Model of Impulse Control*, Harvard Institute of Economic Research Working Papers No. 2112.

[24] Gintis H. (2003). The Hitchhiker's Guide to Altruism: Genes, Culture and the Internalization of Norms, *Journal of Theoretical Biology,* 220, 407-18.

[25] Greenwald A.G. (1980). The Totalitarian Ego. Fabrication and Revision of Personal History, *American Psychologist*, 35, 603-18.

[26] Gul F., Pesendorfer W. (2001). Temptation and Self Control, *Econometrica*, 69, 1403-1436.

[27] Guth W., Yaari M. (1992). An Evolutionary Approach to Explaining Reciprocal Behaviour, in *Simple Strategic Game in Explaining Process Change: Approaches in Evolutionary Economics*, U.Witt (ed.), Ann Arbour, MI: Michigan University Press.

[28] Kamien M.I., Schwartz N.L. (2000). *Dynamic Optimization*, North-Holland.

[29] Laibson D. (1997). Golden Eggs and Hyperbolic Discounting, *Quarterly Journal of Economics*, 112, 443-477.

[30] Lowenstein G., O'Donoghue T. (2004). *Animal Spirits: Affective and Deliberative Processess in Economic Behaviour*, Carnegie Mellon University working paper.

[31] Mancini T. (2001). *Sé e Identità*, Carocci Editore, Roma.

[32] Mancini T. (2006). *Psicologia dell'Identità Etnica. Sé e Appartenenze Culturali*, Roma: Carocci.

[33] Marcia J.E., Waterman A.S., Matteson D.R., Archer S.L., Orlofskyn J.L. (eds.) (1993). *Ego Identity Status. A Handobook for Psychological Research,* New York: Springer-Verlag.

[34] Markus H.R., Wurf E. (1987). The Dynamic Self-Concept: a Social Psychological Perspective, in *Annual Review of Psychology*, 38, 199-377.

[35] Mead G.H. (1934). *Mind, Self and the Society.* Chicago: University of Chicago Press.

[36] O'Donoghue T., Rabin M. (1999). Doing It Now or Later. *American Economic Review,* 89(1), 103-24.

[37] Ok E.A., Vega-Redondo F. (2001). On the Evolution of Individualistic Preferences: an Incomplete Information Scenario, *Journal of Economic Theory,* 97, 231-254.

[38] Rosemberg S. (1988). Self and the Others: Studies in Social Personality and Authobiography, in L. Berkowotz (ed.) *Advances in Experimental and Social Psychology*, New York: Academic Press.

[39] Schelling T.C. (1978). Micromotives and Macrobehaviour , New York: Norton.

[40] Seierstad A., Sydsæter K. (1977). Sufficient Conditions in Optimal Control Theory, *International Economic Review*, 18(2), 367-391.

[41] Strack F., Deutsch R. (2004). Reflective and Impulsive Determinants of Social Behavior, *Personality and Social Psychology Review*, 8(3), 220-247.

[42] Tajfel H.R., Turner J.C. (1979). An Integrative Theory of Intergroup Conflict, in S. Worchel, W.G. Austin (eds.) *The Social Psychology of Intergroup Relations*, Monterey: Brook-Cole.

[43] Turner R.H. (1987). Articulating Self and Social Structure, in K. Yardley, T. Honnes (eds) *Self and Identity: Psychological Perspectives*, Chichester: John Wiley and Sons.

7 Appendix

7.1 Derivation of the dynamic system as a function of a and λ

For expositional easiness, we write the current-value Hamiltonian function and the canonic equations as follows:

$$H = v(a, x, a^e) + \lambda[v_a(a, x, a^e) + g(a, a^m)] \tag{24}$$

$$\dot{a} = v_a(a, x, a^e) + g(a, a^m) \tag{25a}$$

$$\dot{\lambda} = (r - g_a(a, a^m) - v_{aa}(a, a^e))\lambda - v_a(a, x, a^e) \tag{25b}$$

Note that the functions g, g_a only depend on a. The foc for an internal solution are:

$$H_x = v_x(a, x, a^e) + \lambda v_{ax}(a, x, a^e) = 0 \tag{26}$$

thus implicitly defining the optimal value of x^d as a function of the state and costate variables:

$$x^d = F(a, \lambda) \tag{27}$$

The specific value of x^d depends on the specific functional forms. Let us now totally differentiate (27) and (26):

$$dx = F_a da + F_\lambda d\lambda \tag{28}$$

$$H_{xx}dx + H_{ax}da + H_{x\lambda}d\lambda = 0. \tag{29}$$

Since $H_{x\lambda} = v_{ax}$, we can rewrite (29) as:

$$H_{xx}dx + H_{ax}da + v_{ax}d\lambda = 0$$

to obtain

$$dx = -\frac{H_{ax}}{H_{xx}}da - \frac{v_{ax}}{H_{xx}}d\lambda \tag{30}$$

Equating the coefficients of (28) and (30) we get:

$$F_a = -\frac{H_{ax}}{H_{xx}} = -\frac{v_{ax}}{H_{xx}} = F_\lambda. \tag{31}$$

By substituting (27) in system (25) we get:

$$\dot{a} = v_a(a, F(a,\lambda)) + g(a, a^m) \tag{32a}$$

$$\dot{\lambda} = (r - g_a(a, a^m) - v_{aa}(a, a^e))\lambda - v_a(a, F(a,\lambda)). \tag{32b}$$

7.2 Conflictual, polar and extreme dynamic choices

Rewrite the foc (10) as $\frac{n_x(a,x^d)}{m_x(a,x^d)} = 1 - \frac{1}{a+\lambda}$ to show the relation between the marginal rate of substitution $\frac{n_x(a,x^d)}{m_x(a,x^d)}$ and $(a+\lambda)$ (see Fig. 1)

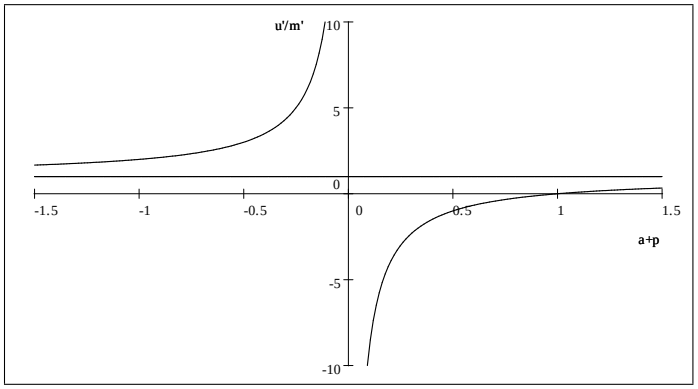

Figure 1: Plot of the marginal rate of substitution as a function of $(a+\lambda)$.

We can distinguish 3 cases:

Compromise choices, which occur when $(a+\lambda) \in (0,1)$. This implies $\frac{n_x(a,x^d)}{m_x(a,x^d)} < 0$, meaning that either

1. $n_x(a, x^d) > 0 > m_x(a, x^d)$, or
2. $m_x(a, x^d) > 0 > n_x(a, x^d)$.

This case can be interpreted by saying that, given a, the optimal dynamic solution x^d requires a mediation between the two sub-utility functions.

Polar choices: occurring in two cases:

1. $(a + \lambda) = 0$, so that $m_x(a, x^d) = 0$ and $x^d = x^m$
2. $(a + \lambda) = 1$, so that $n_x(a, x^d) = 0$ and $x^d = x^n$

Extreme choices, occurring in two cases:

1. $(a+\lambda) < 0$, so that $\frac{n_x(a,x^d)}{m_x(a,x^d)} > 1$. This case requires x^d to stay in a zone where both sub-utility functions are either increasing or decreasing in x, i.e. either $n_x(a, x^d) > m_x(a, x^d) > 0$, or $n_x(a, x^d) < m_x(a, x^d) < 0$
2. $(a+\lambda) > 1$, so that $\frac{n_x(a,x^d)}{m_x(a,x^d)} \in (0, 1)$ This case requires either $m_x(a, x^d) > n_x(a, x^d) > 0$, or $m_x(a, x^d) < n_x(a, x^d) < 0$.