Universitat Autònoma de Barcelona

Essays on Social Networks and Behavioral Economics

Isabel Melguizo López

Universitat Autònoma de Barcelona
Departament d'Economia i d'Història Econòmica

Doctoral Thesis

Essays on Social Networks and Behavioral Economics

Author: Isabel Melguizo López

Supervisor: Miguel Ángel Ballester

Tutor: Inés Macho Stadler

International Doctorate in Economic Analysis

May 2016

Para mi hermano y mis padres

Acknowledgements

I would like to specially thank my supervisor, Miguel Ángel Ballester, for his superb guidance. None of this would have been possible without his support.

Juan de Dios Moreno-Ternero, Pedro Rey-Biel and Tomás Rodríguez Barraquer have been always willing to discuss all the dimensions in which this dissertation could be improved. I strongly acknowledge their contributions.

A part of this project emerged and was consolidated during my visit to the Department of Economics at the University College of London. I am grateful to Antonio Cabrales for offering me the chance to be there.

Interactions with colleagues and professors as well as presentations are key ingredients for growing as a researcher. I thank Tuğçe Çuhadaroğlu for acting as a referee on one part of this dissertation, Jan Zápal for the number of hours he spent reading and discussing every single project, Giacomo De Giorgi for valuable comments and Ezgi Kaya for the time she spent helping me with Matlab. I am grateful to Jorge Alcalde-Unzu and Markus Kinateder for inviting me to present my work at the Departments of Economics at the Public University of Navarra and the University of Navarra, respectively. I thank the Micro-Theory Lab participants for their time and valuable discussions.

At the UAB Economics Department I found an enthusiastic environment to develop my Ph.D. studies. I thank all the faculty members for making me feel at home. My beloved Mercè and Àngels, I do not know what I would have done without your help.

I thank Ana Mauleon and Antonio Miralles and again, Juan de Dios Moreno-Ternero, Pedro Rey-Biel and Jorge Alcade-Unzu for their disposition to form the committee for my thesis defense.

I thank Alberto, Alex, Dilan, Edgardo, Francesco C., Francesco G. and Javi for the shared moments in Barcelona. Special thanks go to Benjamín. I will be always indebted to Ismael, Mila, my aunt Isa and my cousins Javi and Mariana for the time I did not give to them during the preparation of this dissertation.

Financial support from the UAB through PIF Scholarship 912-01-09/2010, from the Spanish Ministry of Economy and Competitiveness through FPI Scholarship BES-2013-06492 and from the Spanish Ministry of Science and Innovation through grants "Consolidated Group-C" ECO2008-04756 and FEDER is gratefully acknowledged.

Contents

Introduction

There are well documented patterns of behavior that individuals exhibit when relating to others or when making economic decisions. For instance, individuals tend to interact disproportionately with others that are similar to them, either in sociodemographic attributes as gender and race, and/or in opinions, attitudes and political views. This phenomenon, known as homophily, is a robust tendency of the way people relate to each other. One of the consequences of individuals behaving this way is that, through a process of social influence, they might reproduce behaviors and thoughts of those with whom they intensively interact. Furthermore, non-cognitive dimensions of personality, as self-control capacity, confidence or perseverance are usually in play, affecting consumption plans or promoting procrastination on tasks.

This dissertation incorporates these patterns of behavior into economic models of social learning and timing decisions related to task performance.

In chapter 1, **Homophily and the Persistence of Disagreement**, we argue how disagreements might perpetuate in society when individuals develop their attitudes by communicating disproportionately with others that are similar to them. Disagreements among individuals might indeed have pernicious consequences, derived from the difficulty in arriving at agreed decisions. In this chapter we consider a dynamic model of attitude formation in which individuals develop their attitudes by incorporating those of others in their social network. Within our framework, individuals are homophilous, that is, the attention they pay to each other is based on whether they possess similar traits. A key and natural feature of our framework is that we allow for the intensity of attention among individuals to vary over time. In particular, the attention to different attributes co-evolves with attitudes, in a way governed by their salience. The salience of an attribute is given by the difference in attitudes between the groups of individuals possessing and lacking it. Since we assume that attributes that exhibit wider differences in attitudes receive more attention, we show that when there is, initially, a unique most salient one, it receives growing attention over time in detriment of the remaining attributes. This process causes the cleavage of the society in two groups of thinking. This situation takes place because individuals redirect their attention towards others similar to them in the initially most salient attribute, sufficiently fast, meaning that this process precludes the complete mixing of attitudes.

Our results can be understood as coming from a process of identity formation. Initially confronted with several attributes according to which individuals might build their personality, they progressively focus in only one of them and, consequently, develop their relationships and attitudes according to it. These results can also be understood as a theory of polarization, since two groups of individuals will be in persistent disagreement.

In chapter 2, **Disagreement in Naïve Models of Attitude Formation: Comparative Statics Results**, we complement the study of chapter 1 by exploring how altering individual behavior in natural directions affects the attitude formation process. It is not difficult to think in real life situations in which individuals do not have certain or well elaborated attitudes about an issue at hand, but these attitudes are noisy. We thus explore the case in which attitudes are subject to shocks. Specifically we consider that individuals' initial attitudes are random draws from a given distribution. We show that disagreement is robust to the introduction of randomness and that is more likely to persist across the attribute for which the distribution of the difference in initial attitudes has the highest mean. We also discuss the case in which individuals relate to each other with different intensities. In fact, as McPherson et al. (2001) document, young and high educated people exhibit lower gender homophily than old and low educated people, respectively. We find that when some individuals exacerbate the attention they pay to similar others, on the basis of the initially most salient attribute, disagreement still persists across it but its magnitude increases and the process of convergence is faster than in the case discussed in chapter 1, when individuals influence each other with the same intensity. We finally explore the question of what are the general conditions that the evolution of homophily has to satisfy for disagreement to persist. In chapter 1 we discuss a particular process in which the evolution of homophily gives raise to persistent disagreement. That is the case because homophily with respect to the initially most salient attribute increases over time in such a way that the convergence of attitudes to a common value is precluded. In contrast, the constant homophily feature in Golub and Jackson (2012) is a key element affecting the speed of convergence to consensus, an outcome that always emerges. Our aim is then to go a step forward in the understanding of what are the homophily patterns that give raise to either persistent disagreement or consensus. In this regard, we find that in general disagreement persists whenever the process by which individuals intensify their relations with others with whom they share the initially most salient attribute, is fast enough. More specifically, there are two forces playing a role: on the one hand individuals pay increasing attention to others on the basis of this attribute but on the other hand, they also always pay a positive amount of attention to everyone else. Disagreement persists if and only if the first force dominates the second.

In chapter 3, **When to Do the Hard Stuff? Dispositions, Motivation and**

the Choice of Difficulties, we discuss the relevance of non-cognitive abilities in the decision of when to face onerous but valuable tasks. For this purpose we consider a dynamic framework in which a decision maker is characterized by the potential with which she can fully execute her non-cognitive abilities. We show that when the decision maker is of low abilities (that is, when she has a low potential) she decides to always face low value easy tasks whereas when the decision maker is of high abilities (that is, when she has a high potential) she decides to always face onerous but valuable tasks. In the latter case the decision maker enjoys higher utility than in the former, in which she decides to always avoid difficulties. We also explore the case in which the execution of non-cognitive abilities is sensitive to outcome achievements, that is, when motivation plays a role. We model motivation by assuming that successes and failures in previous easy tasks affect the decision maker's execution of abilities. We show that in this context the decision maker may decide to jump from low value easy tasks to onerous but valuable tasks at some point in time. Intuitively, being successful in easy tasks motivates the decision maker to do the hard stuff. If indeed individuals behave in this way, there might be policy implications potentially different from the ones derived when considering that human capital accumulation is the main determinant of performance.

Chapter 1

Homophily and the Persistence of Disagreement

1.1 Introduction

Disagreement is an everyday life phenomenon. When in 1987 the American public was confronted with the question of whether the government should guarantee every citizen enough to eat and a place to sleep, 80% of black people agreed whereas only 55% of white people. For around 25 years these percentages have remained almost constant. Disagreements tend to persist, most of the times, over non-factual issues. In fact, differences in attitudes regarding a wide range of topics of ethical and ideological content have persisted among the American public during the aforementioned period.[1]

Despite this evidence, existing models of communication and learning, regardless of whether individuals behave as Bayesian or use rules of thumb, typically lead to consensus. This is the case in DeMarzo et al. (2003), Acemoglu et al. (2010), Golub and Jackson (2010), Golub and Jackson (2012), Smith and Sørensen (2000), Gale and Kariv (2003) and Banerjee and Fudenberg (2004).[2] They are, thus, not suitable for explaining the persistence of disagreement.

The purpose of this paper is to investigate intuitive processes allowing for persistent disagreements. To do so we study the dynamics of attitude formation following DeGroot (1974), a parsimonious and widely used framework in which individuals use the rule of thumb of averaging others' attitudes to develop their own over time. As pointed out by Ellison and Fudenberg (1993), Acemoglu and Ozdaglar (2011) and Golub and Jackson (2012), the computational requirements imposed on agents that behave as Bayesian, updating their (common) priors regarding the true state of the nature according to all relevant information, have placed rules of thumb as a useful and

[1] Detailed information is available at http://www.people-press.org/2012/06/04/section-2-demographics-and-american-values.

[2] We discuss notable exceptions at the end of this section.

powerful alternative for the understanding of learning and communication processes. This reason seems to be borne out by recent evidence supporting, in particular, averaging models as a consistent description of individuals' updating behavior. For instance, experimental results in Chandrasekhar et al. (2012) and Grimm and Mengel (2014) favor a DeGroot procedure over a Bayesian one.[3] The DeGroot procedure allows us to capture the natural idea that individuals usually form and update their attitudes regarding a given issue through own experiences, by observing others' actions and by communicating with others about their attitudes and behavior. That is, learning is social and takes place within the individuals' social network.

But in the canonical DeGroot procedure, which considers a time independent averaging rule, consensus always eventually emerges, under fairly mild conditions. Disagreement only persists in the extreme situation in which there are groups of individuals completely ignoring each other' attitudes.[4] In fact, communication models based on DeGroot procedure, as Golub and Jackson (2010) and Golub and Jackson (2012), work with strongly connected (and time independent) network structures in which individuals incorporate everyone else's attitudes, thus always deriving consensus results. In particular, Golub and Jackson (2012) discuss the effects of homophily, the robust tendency of individuals to associate disproportionately with similar others, on the speed of convergence to consensus, an eventual outcome that is never precluded no matter the level of homophily.[5] In contrast with these papers, we propose a version of DeGroot procedure in which we incorporate the natural idea that the intensity of individual interactions varies over time.[6]

We thus explore a particular mechanism allowing for the co-evolution of homophily in sociodemographic attributes and attitudes. In our approach the type of an individual is defined as a subset of exogenous attributes and we assume that individuals are homophilous with respect to them, that is, common attributes, for instance same gender, are important in determining whether any pair of individuals relate to each other. However, in contrast with the literature, we postulate that the intensity of homophily evolves over time. In particular, it is governed by the salience of attributes. The salience of an attribute is given by the difference in attitudes between individuals possessing and lacking it. The more salient an attribute the higher the attention that, on the basis of it, individuals pay among themselves, that is, the more homophilous towards this attribute individuals are. As a consequence, the lower attention that these individuals pay to others not sharing this attribute with them.

[3] See also Corazzini et al. (2012) and Brandts et al. (2014).

[4] See Jackson (2008), chapter 8, for two characterizations of consensus.

[5] See McPherson et al. (2001) for a survey on homophily. In Golub and Jackson (2012) homophily is technically defined in as the second largest eigenvalue of the matrix of linking densities among types.

[6] See Kossinets and Watts (2006).

There is a large literature in the context of consumer choice supporting the idea that individuals focus in aspects in which their alternatives differ more, that is, in aspects that are salient. For instance, in Bordalo et al. (2013) consumers' purchasing decisions are driven by either the price or the quality of products, depending on which aspect is furthest from prices and qualities of an average bundle.[7] There is also evidence suggesting a negative relationship between differences in attitudes and interactions among individuals. Specifically, Suanet and Van de Vijver (2009) study the relationship between perceived cultural distance, that is, individual reports of discrepancies in attitudes and values between the home and the host culture, and the acculturation of foreign students in Russia. They find a positive (respectively negative) relationship between perceived cultural distance and interactions of foreign students with co-nationals (respectively host nationals). Also, Sole et al. (1975) consider experiments in which individuals have to decide whether to grant help to a stranger. They find a positive relationship between rates of helping and attitude similarity and also how one dissimilar attitude is sufficient to cause significantly lower rates of helping. Finally, Rosenbaum (1986) and Singh and Ho (2000) offer support to the idea that repulsion to others with dissimilar attitudes is the main mechanism shaping homophily.

With this model at hand we answer the following questions:

Q1: Under which conditions does attributes' salience preclude consensus, and therefore, promote the persistence of disagreement?

Q2: How does eventual disagreement look like? In particular, which ones are the types exhibiting different attitudes?

Q3: How does salience relate to the speed of convergence to the eventual situation in which disagreement persist?

Our results are as follows. We find that disagreement persists if and only if there is, initially, a unique attribute for which the difference in average initial attitudes is the highest, that is, a unique most salient attribute. When this is the case, this attribute becomes increasingly salient, receiving growing attention in detriment of the remaining attributes. In other words, the ties among individuals sharing it, will progressively gain strength in detriment of the ties based on the remaining shared attributes. As a result, the society appears eventually divided in two groups of thinking, according to whether individuals possess or lack the initially most salient attribute. Thus, the difference in average eventual attitudes between the groups of individuals possessing and lacking the initially most salient attribute persists while the differences in average

[7] See also Kőszegi and Szeidl (2013) and the references therein.

eventual attitudes associated to the possession and lack of the remaining attributes vanish. As we will see disagreement persists because the aforementioned dynamic is fast enough. By fast enough we mean that the force that drives individuals to develop strong ties with specific individuals dominates the one that pushes them to pay attention to everyone else. Thus, the complete mixing of attitudes is precluded. This process can be understood as one by which individuals construct their identity. That is, initially confronted with several attributes upon which they may build their personality, they progressively focus in only one of them, developing their relations and attitudes according to it. Our results can also be understood as a theory of polarization, since two groups of individuals are in persistent disagreement.

With respect to the properties of disagreement, we find how the difference in average eventual attitudes between the groups of individuals possessing and lacking the initially most salient attribute is a proportion of the difference in average initial attitudes between these two groups.

With respect to the speed of convergence we find that, everything else equal, the higher the difference in average initial attitudes related to the initially most salient attribute or the lower the difference in average initial attitudes related to any other attribute, the higher the magnitude of disagreement and the quicker the convergence to a situation in which it persists.

Our work is related to previous papers discussing disagreement. Specifically, Krause (2000) and Hegselmann and Krause (2002) study disagreement in a model of bounded confidence in which individuals only consider others' attitudes when they are sufficiently close to their own. There are, at least, two differences with their approach. The first one is that while our primary source of attention are individual types as well as their attitudes, they directly focus on similarity in attitudes and do not explicitly model homophily in attributes. The second one is that they assume that the attitudes of the peers finally considered by any individual, matter at the same extent. This is not generally true in our case because homophily depends precisely on types and evolves over time. In Acemoglu et al. (2013) disagreement persists because of the presence of stubborn agents, interpreted as leaders or media sources, that never change their attitudes. We do not model the presence of such agents.

The rest of the paper is organized as follows. Section 1.2 describes the model. Section 1.3 derives the condition for disagreement to persist and provides its properties. Section 1.4 deals with the speed of convergence. Section 1.5 concludes. Section 1.6 contains the technical proofs.

1.2 Preliminaries

Let $I = \{1, 2, \ldots, n\}$ be a finite set of attributes. The type A of an individual is defined by the attributes possessed by this individual, that is, $A \subseteq I$. Thus, there are 2^n types. Let us denote by A^c the complementary set of A. Given two types A and B, we say that they are i-similar whenever attribute i is either present or absent in these two types. Otherwise, we say that they are i-dissimilar. Finally we define $I(AB)$ as all the attributes for which A and B are similar, i.e., $I(AB) = (A \cap B) \cup (A^c \cap B^c)$. Notice that attributes are dichotomous, that is, either a type possesses an attribute or lacks it.[8]

The (column) vector of attitudes at time $t \in \mathbb{Z}_+$ is denoted by $a_t \in [-1, 1]^{2^n}$, where the component relative to type A is a_t^A. The average attitude across all types is denoted $\overline{a}_t$ and the average attitude across all types possessing (respectively lacking) attribute i is denoted $\overline{a}_t[i]$ (respectively $\overline{a}_t[-i]$).[9] Without loss of generality let us normalize the average initial attitude to zero, that is, $\overline{a}_0 = 0$.

Attitudes evolve according to an average-based process similar to DeGroot (1974). Namely, each of the components of a_{t+1} is a weighted average of attitudes in a_t. Let W_t be the $2^n \times 2^n$-matrix of weights describing the updating of attitudes from time t to time $t + 1$. We have that:

$$a_{t+1} = W_t a_t.\text{[10]} \tag{1.1}$$

Notice that every entry of W_t is the weight that type A assigns to type B. Let $w_t^{A,B}$ denote this weight. As in Golub and Jackson (2012), individuals are homophilous, a behavior that can be captured as follows: every attribute i has a non-negative value α_t^i and the weight that type A assigns to type B, is the sum of values of the attributes they share, that is, $w_t^{A,B} = \sum_{i \in I(AB)} \alpha_t^i$. For normalization purposes we set $\sum_i \alpha_t^i = (2^{n-1})^{-1}$. That is the right normalization because a type A is i-similar to exactly 2^{n-1} types. Then, $\sum_B w_t^{A,B} = 2^{n-1} \sum_i \alpha_t^i = 1$. In this paper, we study the case in which homophily, namely, the magnitude of α_t^i, co-evolves with attitudes. In particular, it depends, at every t, on the difference in average attitudes between individuals possessing attribute i and individuals lacking it, that is, $\Delta_t[i] = \overline{a}_t[i] - \overline{a}_t[-i]$ (the salience of attribute i at time t). We assume without loss of generality that the differences in average initial attitudes are non-negative and such that $\Delta_0[1] \geq \Delta_0[2] \geq \cdots \geq \Delta_0[n] \geq 0$.[11]

[8] See Schelling (1969) for a discussion of this assumption. Also, as McPherson et al. (2001) point out, the distinction in terms of social distance appears to be of the type same versus different, and not on any more elaborated forms of stratification.

[9] Formally, $\overline{a}_t = (2^n)^{-1} \sum_A a_t^A$, $\overline{a}_t[i] = (2^{n-1})^{-1} \sum_{A: i \in A} a_t^A$ and $\overline{a}_t[-i] = (2^{n-1})^{-1} \sum_{A: i \notin A} a_t^A$.

[10] With this updating rule, $\overline{a}_0 = 0$ implies that at every time t, $\overline{a}_t = 0$. See section 1.6.

[11] They preserve this order and remain non-negative over time. See step 1 in the proof of the main

We link homophily and salience by the well-known Luce form, that is:

$$\alpha_t^i = \frac{1}{2^{n-1}} \frac{\Delta_t[i]}{\sum_j \Delta_t[j]}. \tag{1.2}$$

We endow this functional form with the following interpretation: the attention that individuals pay to other when they share a given attribute, depends on how big the differences in attitudes associated to this attribute are in relation to the differences associated to the remaining attributes.

The following example illustrates the notation above:

Example 1. Consider the case in which types come from the combination of two attributes. Thus, there are four types, namely, $\{1,2\}$ $\{1\}$, $\{2\}$ and $\{\emptyset\}$. The structure of the 4×4-matrix of weights at an arbitrary time t is:

$$\mathrm{W}_t = \begin{array}{c} \\ \\ \end{array}\begin{array}{cc} \begin{array}{cccc} \{1,2\} & \{1\} & \{2\} & \{\emptyset\} \end{array} & \\ \begin{bmatrix} \alpha_t^1+\alpha_t^2 & \alpha_t^1 & \alpha_t^2 & 0 \\ \alpha_t^1 & \alpha_t^1+\alpha_t^2 & 0 & \alpha_t^2 \\ \alpha_t^2 & 0 & \alpha_t^1+\alpha_t^2 & \alpha_t^1 \\ 0 & \alpha_t^2 & \alpha_t^1 & \alpha_t^1+\alpha_t^2 \end{bmatrix} & \begin{array}{c} \{1,2\} \\ \{1\} \\ \{2\} \\ \{\emptyset\} \end{array} \end{array}$$

To make clear how homophily in exogenous attributes determines the structure of attention, let us consider type $\{2\}$. It is 1-similar to type $\{\emptyset\}$ and 2-similar to type $\{1,2\}$. Thus, it pays a non-negative amount of attention to both types. Also, it pays more attention to these types than to type $\{1\}$, with whom it does not share any attribute. Consider also type $\{1,2\}$. It is 1-similar to type $\{1\}$ and 2-similar to type $\{2\}$. Thus, it pays a non-negative amount of attention attention to them. It also pays zero attention to type $\{\emptyset\}$, with whom it does not share any attribute.

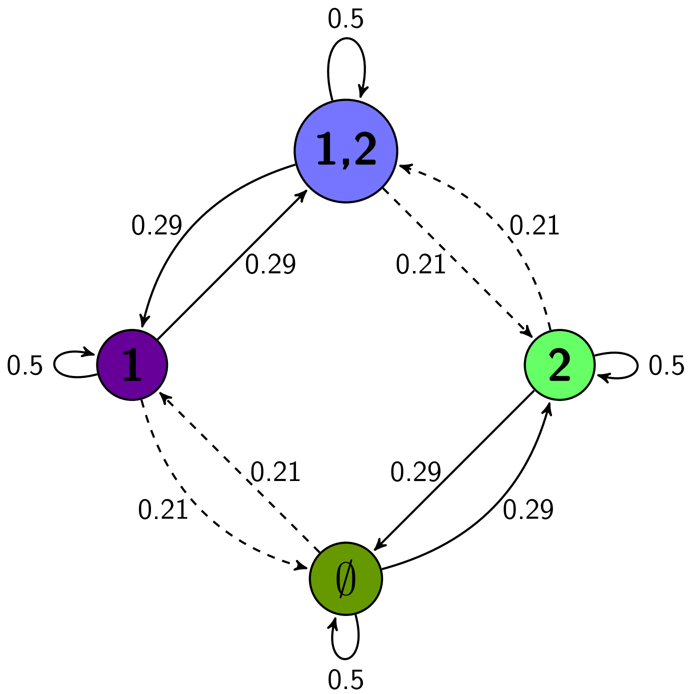
Suppose that initial attitudes are: $a_0^{\{1,2\}} = 0.8$, $a_0^{\{1\}} = 0.2$, $a_0^{\{2\}} = -0.05$ and $a_0^{\{\emptyset\}} = -0.95$. Thus, the differences in average initial attitudes associated to attribute 1 and 2 are $\Delta_0[1] = 0.5(0.8+0.2) - 0.5(-0.05-0.95) = 1$ and $\Delta_0[2] = 0.5(0.8 - 0.05) - 0.5(0.2 - 0.95) = 0.75$, respectively. The initial homophily, driven by salience through expression (1.2), becomes $\alpha_0^1 = 0.29$ and $\alpha_0^2 = 0.21$, for attributes 1 and 2, respectively. The interaction matrix above thus becomes:

$$\mathrm{W}_0 = \begin{array}{cc} \begin{array}{cccc} \{1,2\} & \{1\} & \{2\} & \{\emptyset\} \end{array} & \\ \begin{bmatrix} 0.5 & 0.29 & 0.21 & 0 \\ 0.29 & 0.5 & 0 & 0.21 \\ 0.21 & 0 & 0.5 & 0.29 \\ 0 & 0.21 & 0.29 & 0.5 \end{bmatrix} & \begin{array}{c} \{1,2\} \\ \{1\} \\ \{2\} \\ \{\emptyset\} \end{array} \end{array}$$

Theorem.

In the following figure we depict this interaction structure. For this purpose, let us color types as follows: types possessing attribute 1 are blue and those lacking it are green. Types possessing attribute 2 are white while types lacking it are red. Thus, $\{1,2\}$ is a mixture of blue and white, $\{2\}$ is a mixture of green and white, $\{\emptyset\}$ is a mixture of green and red and $\{1\}$ is a mixture of blue and red:

Figure 1. Depicting initial interactions



We use this structure as a running example in subsequent sections.

1.3 The persistence and properties of disagreement

To analyze under which conditions disagreement persists, notice that expression (1.1) can be solved recursively to get $a_{t+1} = W^T a_0$ where $W^T = \prod_{t=0}^{T} W_{T-t}$. Thus one can express attitudes at an arbitrary point in time t as a function of initial attitudes. Notice that if the matrix describing point-wise interactions, as the one in example 1, was constant over time, consensus will eventually emerge. The reason is that individuals would then be able to incorporate, directly or indirectly, everyone else's attitudes at every point in time. Formally, the (constant) matrix of interactions is strongly connected and aperiodic and thus consensus is guaranteed.[12] In our framework, this is equivalent to establish that all eventual attitudes will be equal to zero. Formally, $a_\infty = \lim_{t \longrightarrow \infty} a_{t+1} = 0$.

[12] See Jackson (2008), chapter 8.

Allowing for the intensity of the attention that individuals pay to each other to vary over time, opens the possibility of persistent disagreement. Clearly, the existence and properties of eventual attitudes can be understood by investigating the existence and properties of the limiting product of time dependent interactions matrices. We denote this limit by W^∞. Formally, $W^\infty = \lim_{T \longrightarrow \infty} W^T$.

In order to state our main result, let us discuss the concept of *Dobrushin* coefficient of ergodicity. Ergodicity coefficients provide information about the extent to which all the rows of a matrix are equal.[13] The Dobrushin coefficient of ergodicity of a matrix M is defined as:

$$\tau(M) = \frac{1}{2} \max_{ij} \sum_k |m_{ik} - m_{jk}|. \tag{1.3}$$

It lies between zero and one and is different from zero if and only if the rows of M are not the same. Now, we present our main result, that describes the form and extent of disagreement in eventual attitudes. It is as follows:

Theorem. *For every configuration of initial attitudes, eventual ones always exist. They exhibit disagreement if and only if attribute* 1 *is, initially, the most salient (that is, if and only if* $\Delta_0[1] > \Delta_0[2]$*). In this case, eventual attitudes are such that, for every type* A*:*

$$|a_\infty^A| = \frac{1}{2}\tau(W^\infty)\Delta_0[1] \tag{1.4}$$

where $\tau(W^\infty) \in (0, 1]$*. Furthermore,* $a_\infty^A > 0$ *if and only if* $1 \in A$*.*

Several aspects merit further attention. First, disagreement is almost the unique outcome of this process.[14] When there is, initially, a unique most salient attribute, it gains increasing attention in detriment of the attention paid to the remaining attributes. Thus, eventual homophily is based upon one, and only one, dimension. Consensus would emerge if and only if there were, at least, two initially most salient attributes. In the extreme case in which all differences in average initial attitudes were equal, all attributes will deserve the same initial homophily, which will be also constant over time. In this case consensus will eventually emerge. Specifically, every attribute i will be receiving always same amount of attention, $\alpha_t^i = (2^{n-1}n)^{-1}$.[15]

[13] See Stachurski (2009) for a reference on the Dobrushin coefficient in the study of economic models with a Markovian structure. See also Ipsen and Selee (2011) and Chatterjee and Seneta (1977) for the study of convergence properties of inhomogeneous Markov chains by means of ergodicity coefficients.

[14] Since differences in average attitudes are real numbers, they are generally different. Also the results remain the same if we consider that initial attitudes are defined over the entire real line instead of belonging to $[-1, 1]$.

[15] When all differences in average initial attitudes are equal to zero, expression (1.2) is not defined. We set $\alpha_t^i = (2^{n-1}n)^{-1}$ in this case. Consensus eventually emerges as well.

Second, disagreement persists between two groups. Specifically, types possessing attribute 1 have the same eventual attitudes and the same happens for types lacking attribute 1. The eventual Attitudes between these two groups are different.

Third, eventual disagreement, measured as the difference in average eventual attitudes between the groups of types possessing and lacking attribute 1, is a proportion of the difference in average initial attitudes between the groups of types possessing and lacking attribute 1. This proportion is exactly given by the ergodicity coefficient of the infinite product of the point-wise matrices of weights.[16] The ergodicity coefficient then characterizes the *distance to consensus in the long-run*. It is important to highlight that this coefficient is fully determined by the initial configurations of attitudes.[17] Also, the difference in average eventual attitudes of types possessing and lacking any attribute different from 1, is zero.[18]

The case with two attributes is pretty informative. In it, the deviation from consensus in the long-run is given by the ratio of the differences in average initial attitudes between attributes 2 and 1. Specifically, $\tau(W^\infty) = 1 - \Delta_0[2]/\Delta_0[1]$. The following example illustrates the results:

Example 2. Consider that $\Delta_0[1] = 1$ and $\Delta_0[2] = 0.75$, as in example 1. The entries in the interaction matrices evolve as follows:

$$W_0 = \begin{bmatrix} 0.5 & 0.29 & 0.21 & 0 \\ 0.29 & 0.5 & 0 & 0.21 \\ 0.21 & 0 & 0.5 & 0.29 \\ 0 & 0.21 & 0.29 & 0.5 \end{bmatrix}, W_1 = \begin{bmatrix} 0.5 & 0.32 & 0.18 & 0 \\ 0.32 & 0.5 & 0 & 0.18 \\ 0.18 & 0 & 0.5 & 0.32 \\ 0 & 0.18 & 0.32 & 0.5 \end{bmatrix}, \ldots, \lim_{t\to\infty} W_t = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}.$$

$$\text{Also,}\ \ W^\infty = \begin{bmatrix} 0.313 & 0.313 & 0.187 & 0.187 \\ 0.313 & 0.313 & 0.187 & 0.187 \\ 0.187 & 0.187 & 0.312 & 0.312 \\ 0.187 & 0.187 & 0.312 & 0.312 \end{bmatrix} \text{and } W^\infty \text{ times } a_0 = \begin{bmatrix} 0.8 \\ 0.2 \\ -0.05 \\ -0.95 \end{bmatrix} \text{ is } a_\infty = \begin{bmatrix} 0.126 \\ 0.126 \\ -0.126 \\ -0.126 \end{bmatrix}.$$

Notice how on one hand, types $\{1, 2\}$ and $\{1\}$ and on the other hand, types $\{2\}$ and $\{\emptyset\}$ hold the same eventual attitudes. Eventual attitudes of these two groups are different. In this case $\tau(W^\infty) = 0.25$.

Fourth, let us denote, for the ease of exposition, $\lambda_t^i = \dfrac{\Delta_t[i]}{\sum_j \Delta_t[j]}$ for every attribute i and at every time t, that is, we make use of expression (1.2) without normalization. It

[16]Let $\overline{a}_\infty[i] = \lim_{t\longrightarrow\infty} \overline{a}_t[i]$ and $\overline{a}_\infty[-i] = \lim_{t\longrightarrow\infty} \overline{a}_t[-i]$. Since there are 2^{n-1} types possessing (respectively lacking) attribute 1, $\overline{a}_\infty[1]-\overline{a}_\infty[-1] = 2^{-1}(\tau(W^\infty)\Delta_0[1]+\tau(W^\infty)\Delta_0[1])=\tau(W^\infty)\Delta_0[1]$.

[17]See step 8 in the proof of the main Theorem.

[18]That is so because within the 2^{n-1} types possessing (respectively lacking) attribute 1, there are 2^{n-2} possessing (respectively lacking) any other attribute $i > 1$, thus the average eventual attitudes of i-similar types are the same and the difference between them cancels out.

follows that the requirement in the main Theorem also holds for the general Luce form, $\gamma_t^i = \frac{\Delta_t[i]^\delta}{\sum_j \Delta_t[j]^\delta}$ when $\delta \in (0, \infty)$.[19] The literature, for instance Chen et al. (1997), interprets δ as a rationality parameter. In our case δ reflects the extent to which the difference in attitudes across attribute 1 is exacerbated. When $\delta \to 0$ this difference become less important than before but disagreement still persists. Intuitively, the magnitude of disagreement happens to be smaller than in the standard case. When exactly $\delta = 0$, the difference in attitudes associated to attribute 1 is as important as the differences associated to any other attribute, regardless of their magnitude. In this case every individual pays always the same attention, $\alpha_t^i = (2^{n-1}n)^{-1}$, to every attribute, thus consensus eventually emerges. When $\delta \to \infty$, the difference in attitudes associated to attribute 1 is increasingly important and the magnitude of disagreement increases with respect to the standard case.

1.3.1 Segregation in interactions and disagreement

As we have discussed, 1-similar types eventually interact exclusively among themselves. They reach this situation by weakening their interactions with 1-dissimilar types.

To summarize this interaction information, we derive here the Spectral Segregation Index proposed by Echenique and Fryer (2007), for attribute i at time t, henceforth SSI_t^i.[20] Being based on the nature of individual interactions, it is particularly suitable in our framework. Other indexes measuring segregation, as the Dissimilarity or the Isolation Index, are based on partitions (census) of a physical unit (a city). In our case individuals are not partitioned into physical units, thus, we do not interpret our interaction process in their terms.[21]

Before stating the result let us stress the fact that interactions within the groups of types possessing and lacking any attribute i, follow the same pattern at every time t. This can be seen using the symmetric interaction matrix in example 1. Interactions among types possessing attribute 1, collapsed in the submatrix composed by $\{1, 2\}$ and $\{1\}$ take the same form as those of types lacking it, collapsed in the submatrix composed by $\{2\}$ and $\{\emptyset\}$. The same is true for attribute 2. Thus, the SSI_t^i describes interactions within both groups. The following result describes the Spectral Index of Segregation for 1-similar types:

Proposition 1. *At every time t, $SSI_t^1 = \frac{1+\lambda_t^1}{2}$. It increases over time, with $SSI_0^1 > \frac{n+1}{2n}$ and $\lim_{t \longrightarrow \infty} SSI_t^1 = 1$.*[22]

[19] In this case $\alpha_t^i = (2^{n-1})^{-1}\gamma_t^i$.

[20] The Spectral Segregation Index has a static nature, we just repeat its computation at every t.

[21] See Echenique and Fryer (2007) for a discussion.

[22] The SSI_t^1 is computed by looking only at interactions among 1-similar types at time t. It is the

Due to our assumptions, the groups of 1-similar types have more intense overall relations than the groups of i-similar types for attributes $i > 1$. That comes from the fact that attribute 1 is always the most salient. Also, interactions among 1-similar types are gradually intensified on the basis of this attribute. We eventually observe the extreme situation in which individuals only interact with others if they are 1-similar. Thus, two disconnected groups, the one composed by types possessing attribute 1 and the one composed by types lacking it, emerge. In this case the segregation of 1-similar types ends up being maximal. In others words, the limiting value of the Spectral Segregation Index is equal to 1.

We can also examine the segregation patterns according to attributes $i > 1$. It turns out that $SSI_t^i = 2^{-1}(1 + \lambda_t^i)$ as well. As attributes $i > 1$ become gradually irrelevant in shaping interactions, segregation according to them is going to decrease over time. Also, the limiting value of the index for attributes $i > 1$ is exactly 0.5.[23] To make this point clearer, recall that eventually two disconnected groups of 1-similar types emerge. If within any of these two groups we focus on the interaction patterns according to any other attribute $i > 1$, we would observe how individuals evenly distribute their unit of attention between i-similar and i-dissimilar types.

Finally, it is worth mentioning the relationship between the segregation of a group of i-similar types and the segregation of the members of this group. By definition, the Spectral Segregation Index is the average of individual segregation indexes. Individual segregation indexes are computed by distributing the overall Spectral Segregation Index among the members of the group. Following Echenique and Fryer (2007), this distribution is done according to the entries of the eigenvector associated to the largest eigenvalue of the matrix describing interactions of i-similar types. In our case this eigenvector is composed by ones. It is then the case that, at every point in time t and for every attribute i, the level segregation of every type is the same, and equal to the overall Spectral Segregation Index. Intuitively, every type pays the same total amount of attention to i-similar types. As a consequence, it also pays the same total amount of attention to i-dissimilar types. In a nutshell, every type segregates its interactions at the same extent and thus equally contributes to the segregation of its group.

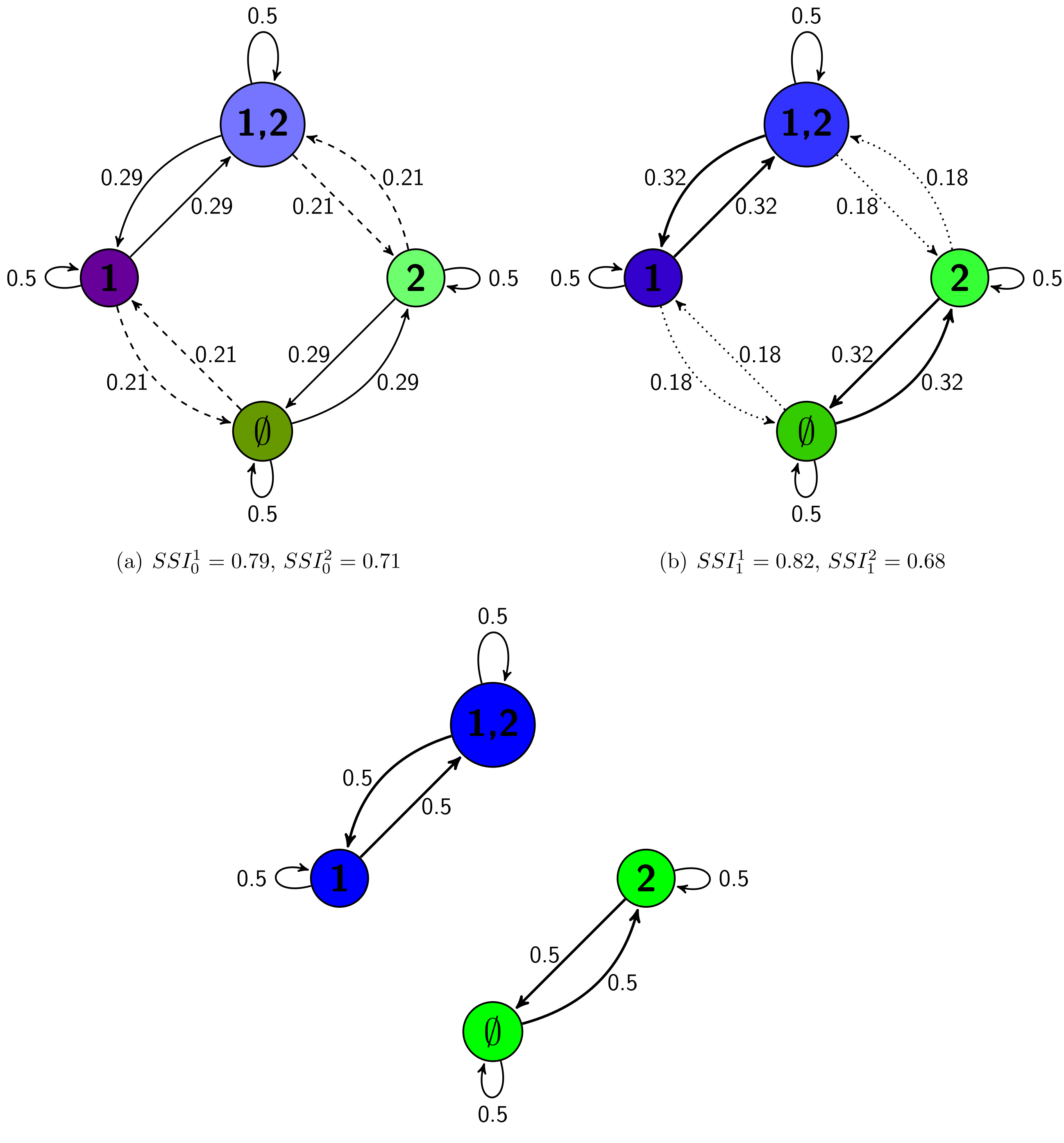
In the following figure we illustrate the evolution of interactions as time goes by and compute the Spectral Segregation Index. Observe how 1-similar types eventually interact exclusively among themselves. Observe also how types possessing (respectively lacking) attribute 2, equally split their unit of attention between themselves and others

largest eigenvalue of the matrix describing these interactions. Also, this result refers to the case in which $\lambda_t^i > 0$ for every attribute i. The results are the same when $\lambda_t^i = 0$ for some/all attributes $i > 1$. We address this case in the proof of this proposition.

[23] Formally, $\lim_{t \longrightarrow \infty} SSI_t^i = 0.5$. Also, $SSI_0^i \leq (n+1)(2n)^{-1}$. As above, the SSI_t^i for every attribute i is computed by looking only at interactions among i-similar types at time t. It is the largest eigenvalue of the matrix describing these interactions.

lacking (respectively possessing) attribute 2.

Figure 2. Segregation in interactions



(a) $SSI_0^1 = 0.79$, $SSI_0^2 = 0.71$

(b) $SSI_1^1 = 0.82$, $SSI_1^2 = 0.68$

(c) $\lim_{t\to\infty} SSI_t^1 = 1$, $\lim_{t\to\infty} SSI_t^2 = 0.5$

Another measure for the intensity of interactions is the so called Network Cohesion, proposed by Cavalcanti et al. (2012). Given a network, represented by a matrix of interactions, Network Cohesion measures how uneven relations are. In other words, how uniform or fragmented a network is. At every time t, Network Cohesion, henceforth C_t, can be computed as one minus the largest eigenvalue of the matrix of interactions W_t. It lies between zero and one, where zero and one represent the lowest and the largest cohesion, respectively. In our framework, λ_t^1, is indeed the largest eigenvalue of the matrix of interactions W_t, thus we have that $C_t = 1 - \lambda_t^1$. Network Cohesion decreases overtime and becomes eventually zero, reflecting the eventual emergence of two disconnected groups of individuals. For instance, in example 2 we have that $\lambda_0^1 = 0.58$, $\lambda_1^1 = 0.64$ and $\lim_{t\to\infty} \lambda_t^1 = 1$, thus $C_0 = 0.42$, $C_1 = 0.36$ and $\lim_{t\to\infty} C_t = 0$.

1.4 Speed of convergence

We focus here on the role of salience in determining the speed of convergence to the eventual disagreement. One reason as to why is relevant to study the speed of convergence is because disagreements might indeed have pernicious consequences. In the presence of a policy intervention aiming to recover consensus, it might be then important to know the timing for its implementation.

As Alesina and Tabellini (1990) point out, discrepancies between policymakers in ideological views about social welfare, specifically regarding the desired composition of government spending in public goods, might cause the accumulation of inefficient levels of public debt. Also, Voss et al. (2006) show how the organizational success of non-profit professional theatres was affected by the divergent views of their leaders regarding the values that should drive the organizations' behavior and Andreoni and Mylovanov (2012) discuss how, among other consequences, disagreement might promote inefficient delays in bargaining. In a broad sense, Friedkin and Johnsen (1999) state that there might be difficulties in arriving at agreed decisions when individuals have fixed discrepant preferences.

The speed of convergence to the eventual disagreement is determined by the relation between the difference in average initial attitudes associated to attribute 1 and the ones associated to the remaining attributes. In other words, the initial relative salience of attribute 1 determines how long it takes for individuals to become sufficiently homophilous with respect to it. Recall that the expression that links homophily based on attribute 1 and the salience of this attribute is given by $\lambda_t^1 = \dfrac{\Delta_t[1]}{\sum_i \Delta_t[i]}$. As previously discussed, eventually 1-similar individuals interact exclusively among themselves which formally means that $\lim_{t\longrightarrow\infty} \lambda_t^1 = 1$. Thus, when we are sufficiently close to this interaction pattern, we can state that we are sufficiently close to the equilibrium in which disagreement persists. It turns out that every time t, λ_t^1 is the second largest eigenvalue of the point-wise matrix of interactions W_t. As deeply discussed in Golub and Jackson (2010) and Golub and Jackson (2012), the second largest eigenvalue of a stochastic matrix plays an important role in the analysis of the speed of convergence.

Our aim in this section is precisely to characterize the time it takes for individuals to become homophilous exclusively with respect to attribute 1, that is, the minimum time it takes for λ_t^1 to be above an $\epsilon > 0$ distance of its limit. For this purpose we formally define this minimum time as:

$$T_\epsilon = min\{t : \lambda_t^1 \geq 1 - \epsilon\}. \tag{1.5}$$

In what follows we describe the properties of T_ϵ, specifically we define its bounds and analyze how it behaves in response to changes in the initial relative salience of attribute 1, that is, to changes in the relation between the difference in average initial attitudes associated to attribute 1 and the ones associated to the remaining attributes. For this purpose, we focus on the case in which all differences in average initial attitudes are strictly positive. We also consider the case in which the relative salience of attribute 1 is modified by altering the differences in average initial attitudes, for just one attribute at a time.[24]

Before stating the result let $r_0^i = \Delta_0[i]/\Delta_0[1]$ for every attribute $i > 1$. This ratio captures the initial relative salience of attribute 1 with respect to any other attribute $i > 1$. The smaller this ratio the more salient attribute 1 is with respect to any other attribute $i > 1$. Let us exceptionally set $\underline{r}_0 = \Delta_0[n]/\Delta_0[1]$ and $\overline{r}_0 = \Delta_0[2]/\Delta_0[1]$. These two ratios represent extreme cases. Specifically, $\underline{r}_0$ considers the difference in average initial attitudes associated to attribute n, which is the smallest one. In contrast, $\overline{r}_0$ considers the difference in average initial attitudes associated to attribute 2, which is the second highest one. Let us set $\overline{\lambda}_t^1 = [1 + (n-1)(\underline{r}_0)^{2^t}]^{-1}$ and define $T_\epsilon^{min} = min\{t : \overline{\lambda}_t^1 \geq 1 - \epsilon\}$ accordingly. Similarly, let $\underline{\lambda}_t^1 = [1 + (n-1)(\overline{r}_0)^{2^t}]^{-1}$ and $T_\epsilon^{max} = min\{t : \underline{\lambda}_t^1 \geq 1 - \epsilon\}$. Notice that both, $\overline{\lambda}_t^1$ and $\underline{\lambda}_t^1$, are constructed from the expression $\lambda_t^1 = [1 + \sum_{i>1}(r_0^i)^{2^t}]^{-1}$, by substituting all differences in average initial attitudes, by the smallest and second highest difference, respectively.[25] We now present the result:

[24]Specifically, for one attribute i, we alter $\Delta_0[i]$ such that $\Delta_0[1] > \Delta_0[2] \geq \cdots \geq \Delta_0[n] \geq 0$ is preserved in order, and in magnitude for differences associated to attributes $j \neq i$. In fact, when we can decrease or increase any $\Delta_0[i]$ by decreasing or increasing, in the same magnitude, initial attitudes of both, the type that possesses all attributes and the type that only possesses attribute i we consider, differences associated to attributes $j \neq i$, keep unaltered. The decrease or increase has to be such that the order above is preserved.

[25]Given the order of initial differences, when $\Delta_0[i] = 0$ for some attribute $i \leq n$ then, $\overline{\lambda}_t^1 = 1$. In this case $T_\epsilon^{min} = 0$. Similarly, when $\Delta_0[2] = 0$ then $\underline{\lambda}_t^1 = 1$. In this case $T_\epsilon^{max} = 0$. Also notice that $\lambda_0^1 = 1$ and thus the equilibrium is reached at $t = 1$.

Proposition 2. *For every configuration of initial attitudes such that disagreement persists, T_ϵ is non-increasing in the initial relative salience of attribute* 1*. Furthermore,* $T_\epsilon \in [T_\epsilon^{min}, T_\epsilon^{max}]$.

It directly follows that, everything else equal, the higher the difference in average initial attitudes associated to attribute 1 the higher the overall attention within the groups of 1-similar types. It is also the case that the lower the difference in average initial attitudes associated to an attribute $i > 1$, the higher the overall attention within the groups of 1-similar types. In particular, attribute 1 becomes relatively more salient than this other, which is now a weaker competitor for attention. In general when attribute 1 is fairly salient, individuals exhibit high homophily with respect to 1-similar others and form completely inward-looking groups relatively fast.

Not only the speed of convergence but the magnitude of disagreement is also sensitive to the aforementioned changes in differences in attitudes. To see this consider the eventual attitudes in expression (1.4) and notice that we can rewrite the ergodicity coefficient as $\tau(W^\infty) = \lim_{T\to\infty} \prod_{t=0}^{T} [1 + \overline{r}_0^{2^t} + ... + \underline{r}_0^{2^t}]^{-1}$. It is direct that the changes in the differences in attitudes decrease the ratios in the denominator, making the elements of this product point-wise higher than before. Thus, the limiting product has to also be higher than before.

It is also worth mentioning how it is enough to focus on the evolution of the homophily value associated to attribute 1 to describe the minimum time of convergence for the system as a whole. The reason is that this homophily value is always further away from 1, its limiting value, than any of the homophily values associated to the remaining attributes is from 0, its limiting value. Then, the time it takes for it to be sufficiently close to one, is at least the same as the time it takes for the remaining homophily values to be sufficiently close to zero.

We finally discuss how the configuration of initial attitudes matters in determining the speed of convergence. Consider the extreme case in which the difference in average initial attitudes associated to attribute 1 is fairly similar to the differences associated to the remaining attributes, that is, $\Delta_0[1] \simeq \Delta_0[2] = \cdots = \Delta_0[n]$. The initial relative salience of attribute 1 is fairly small in this case and it would take a while for individuals to gradually redirect their homophilous behavior towards attribute 1. The time to reach the equilibrium would be considerably high in this case. The other extreme situation is such that the difference in average initial attitudes associated to attribute 1 is, by far, the highest one, for instance, $\Delta_0[1] > \Delta_0[2] = \cdots = \Delta_0[n] \simeq 0$. Being the relative salience of attribute 1 fairly high, individuals would quickly conclude that the possession or lack of this attribute clearly defines two groups in society, or in other words, that this attribute is explanatory for social differences. Thus, it would not take much time for them to become homophilous exclusively with respect to it. The equilibrium will be reached much more faster than before. When the differences in attitudes associated to all attributes $i > 1$ are zero, the equilibrium is reached at $t = 1$.

1.5 Conclusions

On the basis of the observation that disagreement in attitudes is a common phenomenon, we propose a model of attitude evolution able to capture its persistence. In our approach individuals exhibit homophily and the attention they pay to similar others varies over time. Specifically, homophily co-evolves with attitudes governed by the salience of attributes.

We find that disagreement is the long-run outcome of this process if and only if there is a unique attribute that becomes increasingly salient as time goes by. This attribute is precisely the initially most salient one. Thus, eventual homophily is such that individuals only pay attention to others if they are similar to them in that particular attribute. As a product of this behavior, two groups of thinking emerge in the long-run. The time to convergence to this scenario is non-increasing in the initial relative salience of this attribute.

We consider our findings to be related to the phenomenon of unidimensionality in attitudes, a widely discussed topic in political economy. As DeMarzo et al. (2003) point out, there is a strong debate on whether voting records of Congress and Senate members can be explained by a unidimensional liberal-conservative model. There is, in fact, evidence strongly supporting this model. For instance, Poole and Daniels (1985) find that the voting behavior in the U.S. Congress can be mainly explained by a single liberal-conservative dimension. We also consider that our model has a direct application related to the persistence of the gender pay gap. It is sometimes argued that the reason as to why females consistently self-report to be happier at work than males, relies on the fact that they have traditionally held lower labor reward aspirations than males. This phenomenon is known as *The Paradox of Female Happiness*. Two references discussing this paradox and related aspects are Bertrand (2011) and Clark (1997). Divergent aspirations between males and females might be able to explain that part of the gender gap that remains unexplained even after controlling for relevant aspects such as skill levels. Our intuition is that a model of wage setting in which individuals are of both sexes and are endowed with gender biased aspirations, will deliver as a result a gender pay gap, provided that the updating of aspirations takes place with our mechanism. Specifically, females might end up self-selected into low payment jobs, even without discriminatory behavior from the part of employers.

Finally we would like to mention two aspects of the model that we left for future research: first, our model follows a *representative agent approach* in which there is one individual by type. We do not deal with the case in which individuals appear in society in different frequencies. Second, we have assumed that, in determining the intensity of relations individuals sum up the homophily values associated to shared attributes. It will be interesting to investigate the case in which when any pair of individuals share two (or more) attributes i and j, the attention they pay to each other at time t is given by a more general function of α_t^i and α_t^j, and not just its sum.

1.6 Appendix. Proofs

First of all let $\lambda_t^i = \Delta_t[i]/\sum_j \Delta_t[j]$ for every attribute i and at every time t.

Proof of the Theorem. The proof is composed by several steps. In step 1 we show how, at every time t, $\lambda_0^i > 0$ and $\lambda_0^i = 0$ imply that $\lambda_t^i > 0$ and $\lambda_t^i = 0$, respectively. Steps 2-8 analyze disagreement when $\lambda_0^i > 0$ for every attribute i. In particular, Steps 2-5 identify the eigenvalues and eigenvectors of W_t and diagonalize it. Step 6 deals with the existence of the limiting product of point-wise stochastic matrices, that is, W^∞. Step 7 provides its form. Step 8 establishes the necessary and sufficient condition for disagreement to persist, qualifying it. Finally, step 9 describes the case in which $\lambda_0^i = 0$ for some/all attributes $i > 1$.

Step 1. We prove that $\lambda_0^i > 0$ implies that $\lambda_t^i > 0$ and $\lambda_0^i = 0$ implies that $\lambda_t^i = 0$, at every time t. We proceed by decomposing $\Delta_t[i] = (2^{n-1})^{-1}\left[\sum_{A:i\in A} a_t^A - \sum_{A:i\notin A} a_t^A\right]$. Consider a type A such that $i \in A$. By (1.1), $a_t^A = \sum_B w_{t-1}^{A,B} a_{t-1}^B$. Since $w_{t-1}^{A,B} = (2^{n-1})^{-1}\sum_{i\in I(AB)} \lambda_{t-1}^i$, then:

$$a_t^A = \sum_B w_{t-1}^{A,B} a_{t-1}^B = \frac{1}{2^{n-1}} \sum_{B:i\in B} \lambda_{t-1}^i a_{t-1}^B + \frac{1}{2^{n-1}} \sum_{j\neq i} \lambda_{t-1}^j \sum_{B:j\in I(AB)} a_{t-1}^B.$$

Since there are 2^{n-1} types A possessing attribute i, $\sum_{A:i\in A} a_t^A = \sum_{B:i\in B} \lambda_{t-1}^i a_{t-1}^B + \sum_{j\neq i} \lambda_{t-1}^j \sum_{B:j\in I(AB)} a_{t-1}^B$. By a similar reasoning, for types A such that $i \notin A$, $\sum_{A:i\notin A} a_t^A = \sum_{B:i\notin B} \lambda_{t-1}^i a_{t-1}^B + \sum_{j\neq i} \lambda_{t-1}^j \sum_{B:j\in I(AB)} a_{t-1}^B$. Therefore:

$$\frac{1}{2^{n-1}} \sum_{A:i\in A} a_t^A - \frac{1}{2^{n-1}} \sum_{A:i\notin A} a_t^A = \frac{1}{2^{n-1}} \sum_{B:i\in B} \lambda_{t-1}^i a_{t-1}^B - \frac{1}{2^{n-1}} \sum_{B:i\notin B} \lambda_{t-1}^i a_{t-1}^B$$

or equivalently, $\Delta_t[i] = \lambda_{t-1}^i \Delta_{t-1}[i]$.[26] From the definition of λ_t^i, it follows that at every t, $\Delta_t[i] \geq 0$ implies that $\lambda_t^i \geq 0$. Also, $\Delta_t[i] \geq 0$ if and only if $\lambda_{t-1}^i \geq 0$ and $\Delta_{t-1}[i] \geq 0$. With these two observations we conclude that $\Delta_0[i] > 0$ and $\lambda_0^i > 0$ imply that at every time t, $\Delta_t[i] > 0$ and $\lambda_t^i > 0$, respectively. Also $\Delta_0[i] = 0$ and $\lambda_0^i = 0$ imply that at every time t, $\Delta_t[i] = 0$ and $\lambda_0^i = 0$, respectively.[27]

[26] As this expression holds at every t, we recursively write $\Delta_t[i] = \prod_{s=0}^{t-1} \lambda_s^i \Delta_0[i]$.

[27] Also from the definition of λ_t^i it follows that at every t, $\Delta_t[1] \geq \Delta_t[2] \geq \cdots \geq \Delta_t[n] \geq 0$ implies that $\lambda_t^1 \geq \lambda_t^2 \geq \cdots \geq \lambda_t^n \geq 0$. Additionally, $\Delta_t[1] \geq \Delta_t[2] \geq \cdots \geq \Delta_t[n] \geq 0$ if and only if $\lambda_{t-1}^1 \Delta_{t-1}[1] \geq \lambda_{t-1}^2 \Delta_{t-1}[2] \geq \cdots \geq \lambda_{t-1}^n \Delta_{t-1}[n] \geq 0$. Since by assumption $\Delta_0[1] \geq \Delta_0[2] \geq \cdots \geq \Delta_0[n] \geq 0$ then, at every t, $\Delta_t[1] \geq \Delta_t[2] \geq \cdots \geq \Delta_t[n] \geq 0$ and $\lambda_t^1 \geq \lambda_t^2 \geq \cdots \geq \lambda_t^n \geq 0$ hold. This also implies that $\prod_{t=0}^T \lambda_t^1 > \prod_{t=0}^T \lambda_t^2 \geq \cdots \geq \prod_{t=0}^T \lambda_t^n \geq 0$.

Step 2. At every time t, 1 is an eigenvalue of W_t, with right-eigenvector u of size $2^n \times 1$, where u has all components equal to 1. This directly follows from the stochasticity of W_t. Notice that u is time independent. We thus omit the time subscript.

Step 3. At every time t and for every attribute i, λ_t^i is an eigenvalue of W_t, with right-eigenvector u^i of size $2^n \times 1$, where u^i has the following form: the component of u^i associated to type A is equal to 1 if $i \in A$ and equal to -1 otherwise. We prove that by showing that the pair (λ_t^i, u^i) satisfies the eigenvalue equation, $W_t u^i = \lambda_t^i u^i$. Consider an attribute i and an arbitrary type A. Suppose first that $i \in A$. Notice that there are exactly 2^{n-1} types B possessing attribute i. Also, notice that for every $j \neq i$, there are exactly 2^{n-2} types B possessing attribute i that are j-similar to A and 2^{n-2} types B lacking attribute i that are j-similar to A. Therefore, the row in W_t corresponding to type A, multiplied by u^i, is equal to:

$$\sum_{B:i\in B} w_t^{A,B} - \sum_{B:i\notin B} w_t^{A,B} = \frac{2^{n-1}\lambda_t^i + 2^{n-2}\sum_{j\neq i}^{n}\lambda_t^j - 2^{n-2}\sum_{j\neq i}^{n}\lambda_t^j}{2^{n-1}} = \lambda_t^i.$$

Since every type A is such that $i \in A$, the RHS of the eigenvalue equation also equals λ_t^i. Thus, we conclude that (λ_t^i, u^i) is a pair of eigenvalue and right-eigenvector of W_t. The proof for the case in which A is such that $i \notin A$ is analogous and hence omitted. As in step 2, the eigenvectors u^i corresponding to every λ_t^i are also time independent.

Step 4. At every time t, the remaining eigenvalues of W_t are zero. Consider any type B such that $|B| \geq 2$. We start by proving that for every type A, $w_t^{A,B} = \sum_{i\in B} w_t^{A,\{i\}} - [|B|-1]\, w_t^{A,\emptyset}$. By doing so, we are proving that the column vector of weights associated to type B is a linear combination of the column vectors of weights associated to types containing at most one attribute and hence, there are at most $n+1$ independent columns in W_t. Notice that getting rid of the normalization $1/2^{n-1}$, we are left with:

$$\sum_{i\in B} w_t^{A,\{i\}} = \sum_{i\in B\cap A}(\lambda_t^i + \sum_{j\in A^c}\lambda_t^j) + \sum_{i\in B\cap A^c}\sum_{j\in A^c, j\neq i}\lambda_t^j$$

and that this is equivalent to:

$$\sum_{i\in B\cap A}(\lambda_t^i + \sum_{j\in A^c}\lambda_t^j) + \sum_{i\in B\cap A^c}(\sum_{j\in A^c}\lambda_t^j - \lambda_t^i) = \sum_{i\in B\cap A}\lambda_t^i + \sum_{i\in B}\sum_{j\in A^c}\lambda_t^j - \sum_{i\in B\cap A^c}\lambda_t^i. \quad (1.6)$$

Second, notice that:

$$(|B|-1)w_t^{A,\emptyset} = (|B|-1)\sum_{j\in A^c}\lambda_t^j. \quad (1.7)$$

Thus, the difference between expression (1.6) and (1.7) is equal to $\sum_{i\in B\cap A}\lambda_t^i + \sum_{j\in A^c}\lambda_t^j - \sum_{i\in B\cap A^c}\lambda_t^i$. This expression can be rewritten as $\sum_{i\in B\cap A}\lambda_t^i + \sum_{j\in I\cap A^c}\lambda_t^j - \sum_{i\in (I\setminus B^c)\cap A^c}\lambda_t^i = \sum_{i\in B\cap A}\lambda_t^i + \sum_{i\in B^c\cap A^c}\lambda_t^i$. This is equivalent to $w_t^{A,B} = \sum_{i\in I(AB)}\lambda_t^i$ where $I(AB) = (B\cap A)\cup(B^c\cap A^c)$ as defined in section 1.2. Thus, $rank(W_t) \leq n+1$.

Recall that the rank of a matrix is equal to the number of non-zero eigenvalues. Since steps 2 and 3 already identified $n+1$ of them, indeed $rank(W_t) = n+1$. Thus, the rest of the $2^n - (n+1)$ eigenvalues are zero.

Step 5. We prove here that W_t is always diagonalizable and provide its form. From symmetry of W_t there is an orthogonal diagonalization $W_t = U\Lambda_t U'$, where U is an orthonormal basis. Orthonormal eigenvectors have unitary euclidean norm and are orthogonal to each other. Therefore, for the zero eigenvalues, there exist eigenvectors u^0 with $\|u^0\| = 1$, orthogonal to each other and to both, $u/\|u\|$ and every $u^i/\|u^i\|$, where $\|u\| = 2^{n/2}$ and $\|u^i\| = 2^{n/2}$, for every i. Since by steps 2 and 3 u and every u^i are time independent, every u^0 is also time independent. Now, fix the following order of eigenvalues: first eigenvalue 1, afterwards eigenvalues λ_t^i, by type, and finally the zero eigenvalues in a fixed order. Then $U = \left[\frac{u}{\|u\|}\ \frac{u^i}{\|u^1\|}\cdots\frac{u^n}{\|u^n\|}\ u^0\ldots u^0\right]$, and the diagonal matrix of eigenvalues at time t is:

$$\Lambda_t = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \lambda_t^1 & \cdots & \cdots & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & \lambda_t^n & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Since at every time t the matrix W_t is diagonalizable over the same eigenspace, hence $W^T = U\Lambda^T U'$ where $\Lambda^T = \prod_{t=0}^T \Lambda_t$ with diagonal entries: 1, $\prod_{t=0}^T \lambda_t^i$ for every attribute i and zeros.

Step 6. Here we deal with the existence of W^∞ and a_∞. By step 5, $W^\infty = U\lim_{T\longrightarrow\infty}\Lambda^T U'$, provided that the RHS of this expression exists. We confirm here that this is, in fact, the case. In computing $\lim_{T\longrightarrow\infty}\Lambda^T$ we focus on the non-zero diagonal entries of Λ^T. Eigenvalue 1 is constant over time, thus its limiting product is 1. Since at every time t $\lambda_t^i \in (0,1)$ for every i, $\prod_{t=0}^\infty \lambda_t^i$ exists in $[0,1)$. Thus, $U\lim_{T\longrightarrow\infty}\Lambda^T U'$ exists and defines both, W^∞ and $a_\infty = W^\infty a_0$, for every a_0.

Step 7. We provide here the specific form of W^∞. Suppose that $\Delta_0[1] > \Delta_0[2]$. Consider attribute 1 first. Let $r_t^i = \Delta_t[i]/\Delta_t[1]$ for every attribute i and at every time t. We then rewrite $\lambda_t^1 = (\Delta_t[1])(\Delta_t[1] + \sum_{i>1}\Delta_t[i])^{-1} = [1 + \sum_{i>1} r_t^i]^{-1}$. By step 1, $r_t^i = \lambda_{t-1}^i\Delta_{t-1}[i]/\lambda_{t-1}^1\Delta_{t-1}[1]$. From the expression of λ_t^i it follows that $\lambda_{t-1}^i/\lambda_{t-1}^1 = \Delta_{t-1}[i]/\Delta_{t-1}[1] = r_{t-1}^i$. Thus, $r_t^i = (r_{t-1}^i)^2$ and recursively we get that $r_t^i = (r_0^i)^{2^t}$.

Thus, $\lambda_t^1 = [1+\sum_{i>1}(r_0^i)^{2^t}]^{-1}$. It is important to notice that, $0 < r_0^i < 1$ for attributes $i > 1$. It then follows that $\lim_{t\longrightarrow\infty} \lambda_t^1 = 1$. This opens the possibility for $\prod_{t=0}^{\infty} \lambda_t^1 \neq 0$. We prove that this is indeed the case by equivalently stating that $\sum_{t=0}^{\infty} log(\lambda_t^1)$ exists.

In order to do it, we consider r_0^2, the highest ratio smaller than one, and construct a new homophily value as follows: we replace r_0^i, for attributes $i > 1$, with r_0^2 in λ_t^1. Specifically, we have that $\underline{\lambda}_t^1 = [1+(n-1)(r_0^2)^{2^t}]^{-1}$. Since $r_0^2 \geq r_0^i$ for every $i > 1$, then $\lambda_t^1 \geq \underline{\lambda}_t^1$ at every time t. We prove that $\sum_{t=0}^{\infty} log(\underline{\lambda}_t^1)$ exists, so does $\sum_{t=0}^{\infty} log(\lambda_t^1)$, by comparison. We proceed by testing the absolute convergence (and hence the convergence) of $\sum_{t=0}^{\infty} log(\underline{\lambda}_t^1)$, using the ratio test. It is well known that an adaptation of the L'Hopital rule can be used to find limits of sequences. We thus define $f(x)$ and $g(x)$ as functions of a real variable x and $\{s_t\}$ such that at every t, $s_t = f(t)/g(t)$. Then, we evaluate $\lim_{x\rightarrow\infty} f(x)/g(x) = \lim_{x\rightarrow\infty} \dfrac{log(1+(n-1)(r_0^2)^{2^{x+1}})}{log(1+(n-1)(r_0^2)^{2^x})}$. Since $0 < r_0^2 < 1$, this limit is indeterminate. By L'Hopital $\lim_{x\rightarrow\infty} f(x)/g(x) = \lim_{x\rightarrow\infty} f'(x)/g'(x) = \dfrac{2(r_0^2)^{2^x}(1+(n-1)(r_0^2)^{2^x})}{(1+(n-1)(r_0^2)^{2^{x+1}})} = 0$. Thus, $\lim_{t\longrightarrow\infty} s_t = \lim_{x\rightarrow\infty} f(x)/g(x) = 0$. This implies that $\sum_{t=0}^{\infty} |log(\underline{\lambda}_t^1)|$ exists. Since at every t, $\lambda_t^1 \geq \underline{\lambda}_t^1$, then $|log(\lambda_t^1)| \leq |log(\underline{\lambda}_t^1)|$. Thus, by comparison $\sum_{t=0}^{\infty} |log(\lambda_t^1)|$ exists, so does $\sum_{t=0}^{\infty} log(\lambda_t^1)$.

Consider now attributes $i > 1$. For a given $i > 1$, let j denote attributes other than it and let $r_t^j = \Delta_t[j]/\Delta_t[i]$. Then $\lambda_t^i = [1+\sum_{j\neq i}(r_0^j)^{2^t}]^{-1}$. Notice that $r_0^1 > 1$. Then $\lim_{t\longrightarrow\infty} \lambda_t^i = 0$ and $\prod_{t=0}^{\infty} \lambda_t^i = 0$ for attributes $i > 1$. Summing up we have that $\prod_{t=0}^{\infty} \lambda_t^1 = \mu^1$ with $\mu^1 \in (0,1)$ and $\prod_{t=0}^{\infty} \lambda_t^i = 0$ for $i > 1$. Under this scenario:

$$\lim_{T\longrightarrow\infty} \Lambda^T = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \mu^1 & \vdots & \vdots & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \vdots & \vdots & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad U \lim_{T\longrightarrow\infty} \Lambda^T = \frac{1}{2^{n/2}} \begin{bmatrix} 1 & \mu^1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & 0 \\ 1 & \mu^1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 1 & -\mu^1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & 0 \\ 1 & -\mu^1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and thus,

$$W^{\infty} = U \lim_{T\longrightarrow\infty} \Lambda^T U' = \frac{1}{2^n} \begin{bmatrix} 1+\mu^1 & \cdots & 1+\mu^1 & 1-\mu^1 & \cdots & 1-\mu^1 \\ 1+\mu^1 & \cdots & 1+\mu^1 & 1-\mu^1 & \cdots & 1-\mu^1 \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 1+\mu^1 & \cdots & 1+\mu^1 & 1-\mu^1 & \cdots & 1-\mu^1 \\ 1+\mu^1 & \cdots & 1+\mu^1 & 1-\mu^1 & \cdots & 1-\mu^1 \\ 1-\mu^1 & \cdots & 1-\mu^1 & 1+\mu^1 & \cdots & 1+\mu^1 \\ 1-\mu^1 & \cdots & 1-\mu^1 & 1+\mu^1 & \cdots & 1+\mu^1 \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 1-\mu^1 & \cdots & 1-\mu^1 & 1+\mu^1 & \cdots & 1+\mu^1 \\ 1-\mu^1 & \cdots & 1-\mu^1 & 1+\mu^1 & \cdots & 1+\mu^1 \end{bmatrix}$$

.

Notice that since eigenvectors u^0, associated to the zero eigenvalues, occupy the last columns (respectively rows) of U (respectively U'), they are not involved in these products.

Suppose now that $\Delta_0[1] = \Delta_0[i]$ for some attributes $i > 1$. Let e be the number of attributes $i > 1$ such that $\Delta_0[1]$ and $\Delta_0[i]$ are equal. Then $\lim_{t\longrightarrow\infty} \lambda_t^1 = [e+1]^{-1} \neq 1$, This implies that $\prod_{t=0}^{\infty} \lambda_t^1 = 0$. By the same reasoning this is also the case for attributes $i > 1$ such that $\Delta_0[1] = \Delta_0[i]$. For attributes $i > 1$ such that $\Delta_0[1] > \Delta_0[i]$, then $\lim_{t\longrightarrow\infty} \lambda_t^i = 0$ by a similar arguments as above and thus $\prod_{t=0}^{\infty} \lambda_t^i = 0$ for them. Under this scenario, every entry in W^∞ is $(2^n)^{-1}$.

Before concluding let us consider the general case in which $\gamma_t^i = \Delta_t[i]^\delta / \sum_j \Delta_t[j]^\delta$ with $\delta \in (0, \infty)$. We can rewrite $\gamma_t^1 = [1 + \sum_{i>1} (r_0^i)^{\delta(\delta+1)^t}]^{-1}$ in this case. Notice that $\lim_{t\to\infty} \gamma_t^1 = 1$ and thus, $\lim_{t\to\infty} \gamma_t^i = 0$ for $i > 1$. We study the convergence of $\sum_{t=0}^{\infty} log(\gamma_t^1)$ using the same reasoning as before, where now $\underline{\gamma}_t^1 = [1 + (n-1)(r_0^2)^{\delta(\delta+1)^t}]^{-1}$. Using similar algebra and reasoning as above we conclude that $\lim_{x\to\infty} f'(x)/g'(x) = \dfrac{2(r_0^2)^{\delta^2(\delta+1)^x}(1+(n-1)(r_0^2)^{\delta^2(\delta+1)^x})}{(1+(n-1)(r_0^2)^{\delta^2(\delta+1)^{x+1}})}$=0 for $\delta \neq 0$. It is then the case that $\sum_{t=0}^{\infty} log(\gamma_t^1)$ converges, meaning that disagreement persists. When $\Delta_0[1] = \Delta_0[i]$ for some attribute(s) $i > 1$ consensus emerges as above.

Step 8. We establish here the necessary and sufficient condition for disagreement to persist. We also qualify disagreement. Recall that $\overline{a}_0 = 0$. The eventual attitude of a type A is the result of multiplying its corresponding row in W^∞ times the column vector of initial attitudes. Consider first that $\Delta_0[1] > \Delta_0[2]$. Then W^∞ is the one derived in step 7. For the first 2^{n-1} rows of W^∞, corresponding to types A such that $1 \in A$, we thus have that $a_\infty^A = 2^{-1}\mu^1 \left[\dfrac{1}{2^{n-1}} \sum_{A:i\in A} a_0^A - \dfrac{1}{2^{n-1}} \sum_{A:i\notin A} a_0^A\right] = 2^{-1}\mu^1\Delta_0[1]$. For the subsequent 2^{n-1} rows corresponding to types A such that $1 \notin A$, $a_\infty^A = -\dfrac{1}{2}\mu^1\Delta_0[1]$. Thus, in general, for every type A, $|a_\infty^A| = 2^{-1}\mu^1\Delta_0[1]$ and eventual attitudes are positive if and only if A is such that $1 \in A$. That is, disagreement persists.

We are left to prove that $\tau(W^\infty) = \mu^1$. Consider expression (1.3). Fixing any column in W^∞, the maximum distance between any two rows is $\mu^1/2^{n-1}$, which summing across the 2^n columns and dividing by 2 yields μ^1. Finally, since $\mu^1 = \prod_{t=0}^{\infty} \lambda_t^i = \lim_{T\to\infty} \prod_{t=0}^{T} \left[1 + \sum_{i>1} (\Delta_0[i](\Delta_0[1])^{-1})^{2^t}\right]^{-1}$, we have that, $|a_\infty^A| = 2^{-1}\tau(W^\infty)\Delta_0[1]$.

Consider now that $\Delta_0[1] = \Delta_0[i]$ for some attributes $i > 1$. By step 7, every entry of W^∞ is $(2^n)^{-1}$. In this case $a_\infty^A = 0$ for every type A. That is, consensus eventually emerges.

We then conclude that disagreement persists if and only if attribute 1 is, initially, the unique most salient attribute.

Step 9. We consider the case in which $\lambda_0^i = 0$ for some/all attributes $i > 1$. Step 1 relies on the linearity of the updating process. Thus, it still holds. Since at every t, W_t remains stochastic, step 2 holds. For the attributes i such that $\lambda_t^i > 0$, the statement in step 3 hold as well. Step 4 holds with the difference that now there are $2^n - (n+1-N)$ zero eigenvalues, where N is the number of attributes i such that $\lambda_t^i = 0$. In the extreme case in which $\lambda_t^i = 0$ for every attribute $i > 1$, the column corresponding to the empty type and the $n-1$ columns corresponding to the singleton types with attributes different from 1, are the same. In such a case $N = n-1$ and there are 2 independent columns. The eigenvalues different from zero at every t are 1 because of stochasticity and $\lambda_t^1 = 1$. Since at every t, W_t remains symmetric, step 5 holds. Step 6 deals with the existence of W^∞, which is based on the existence of the limiting product of non-zero eigenvalues. It also goes through. Since the form of W^∞ depends only on whether $\Delta_0[1] > \Delta_0[2]$, despite of λ_t^i being 0 for some/all attributes $i > 1$, step 7 holds. Finally step 8, that establishes the necessary and sufficient condition for disagreement to persist, qualifying it, also holds. Notice that when $\lambda_0^i = 0$ for all attributes $i > 1$ then $W^\infty = W_0$. Also, $\mu^1 = 1$ and the equilibrium is reached at $t = 1$.

■

Proof of Proposition 1. We compute here the Spectral Index of Segregation at every time t. For this purpose we directly follow Echenique and Fryer (2007). Before proceeding recall that by step 1 in the proof of the main Theorem, positive (respectively zero) homophily values remain positive (respectively zero) all along the process. Recall also that $\sum_i \lambda_t^i = 1$ at every time t. Consider first the case in which for every attribute i, $\lambda_t^i > 0$.

Consider only types possessing attribute 1. Denote the matrix of their interactions by $\mathbf{1}_t$. Since all types have attribute 1 in common, they pay a positive amount of attention to each other, thus $\mathbf{1}_t$ has only one connected component composed by all individuals in $\mathbf{1}_t$. We now compute the largest eigenvalue of $\mathbf{1}_t$. Our claim is that $\lambda_t = \lambda_t^1 + 2^{-1}\sum_{j\neq i}^n \lambda_t^j$, with associated time independent right-eigenvector u of size $2^{n-1} \times 1$, where u is composed by ones, is the largest eigenvalue of $\mathbf{1}_t$. We first prove that (λ_t, u) is a pair of eigenvalue and right-eigenvector of $\mathbf{1}_t$. Second, we argue that λ_t is the largest eigenvalue of $\mathbf{1}_t$.

First, notice that every type A shares attribute 1 with 2^{n-1} types. It also shares the rest of attributes with 2^{n-2} types. Thus, any row of $\mathbf{1}_t$ by u reads $(2^{n-1}\lambda_t^1 + 2^{n-2}\sum_{j\neq i}^n \lambda_t^j)(2^{n-1})^{-1}$. This is equivalent to $\lambda_t \times 1$. Therefore, the eigenvalue equation is satisfied and (λ_t, u) is a pair of eigenvalue and (column) eigenvector of $\mathbf{1}_t$. Second, by Perron-Froebenius Theorem, being $\mathbf{1}_t$ a positive matrix, it has a unique largest eigenvalue, which is strictly positive (that is, the spectral radius of $\mathbf{1}_t$). It is bounded above by the maximum sum of the entries of a row in $\mathbf{1}_t$ (see Meyer (2000), chapter 8). Notice that every row of $\mathbf{1}_t$ sums up to the same value, which is precisely λ_t. Suppose that there is other positive real eigenvalue, different than λ_t, which is the largest. Then it has to be also larger than the maximum sum of the entries of a row in $\mathbf{1}_t$, contradicting the Perron-Froebenius Theorem. Then, λ_t has to be the largest eigenvalue. We rewrite it as $\lambda_t = \lambda_t^1 + 2^{-1}(1 - \lambda_t^1) = 2^{-1}(1 + \lambda_t^1)$. Let us denote $SSI_t^1 = \lambda_t$. Finally, it directly follows that λ_t increases with λ_t^1. Since $\lim_{t\longrightarrow\infty} \lambda_t^1 = 1$ then $\lim_{t\longrightarrow\infty} SSI_t^1 = 1$ as well. Also, if every attribute i was initially equally salient, then $\lambda_0^i = 1/n$ for each of them. Since attribute 1 is the initially most salient, it has to be that $\lambda_0^i > 1/n$. Thus, $SSI_0^1 > (n+1)(2n)^{-1}$. Notice that the analysis is exactly the same when we consider interactions of types lacking attribute 1. In fact, the matrix of interactions is exactly $\mathbf{1}_t$. Also, in computing the SSI_t^i for attributes $i > 1$, we follow similar arguments. Thus, we omit the proofs.

Consider now the case in which for attribute 1, $\lambda_t^1 > 0$ and for some/all attributes $i > 1$, $\lambda_t^i = 0$.[28] We prove here that when for an attribute i, $\lambda_t^i = 0$ then the SSI_t^i is, at every t, equal to one half.[29] Given the evolution of the homophily values, as described in the proof of Theorem 1, this is also its limiting value. Recall that, by step 1 in the proof of Theorem 1, when for an attribute i such that $2 \leq i \leq n$, $\lambda_t^i = 0$, this implies that $\lambda_t^j = 0$ for all attributes $j > i$. Let us focus on types possessing attribute i. The analysis is exactly the same when we consider interactions of types lacking attribute i. Two cases arise:

C.1. Suppose that for every attribute j such that $1 < j < i$, then $\lambda_t^j = 0$, then interactions among types possessing attribute i are defined by two connected components, based on the lack or possession of attribute 1. The matrices defining these two connected components are the same and have all their entries positive. One of the matrices has 2^{n-2} types possessing attribute 1 and the other has 2^{n-2} types lacking it. The analysis within each matrix is exactly the same as before. In each of them, the sum of every row is $2^{n-2}(2^{n-1})^{-1}\lambda_t^1 = 0.5$. Thus, within each component, SSI_t^i equals to one half at every time t. Thus, the average of the SSI_t^i of each component is also equal to one half.

C.2. Suppose that for some/all attributes j such that $1 < j < i$, $\lambda_t^j > 0$. In this case there is only one connected component. The reason is that types possessing (respectively lacking) attribute 1 are always connected among themselves and these two groups are connected between them since both contain types that are similar in attributes $j < i$, with $\lambda_t^j > 0$. The sum of the entries of every row of the matrix of interactions is $2^{n-2}(\lambda_t^1 + \sum_{j\neq i} \lambda_t^j)(2^{n-1})^{-1} = 0.5$. Thus, the index is equal to one half at every time t.[30] ■

[28] Recall that when all differences in average initial attitudes are equal, either positive or zero, then $\lambda_t^i = 1/n$ for every i and at every t. Then, $SSI_t^i = (n+1)(2n)^{-1}$ for every attribute i and at every t. See the proof of the main Theorem.

[29] Notice that when computing the SSI_t^i for an attribute i such that $\lambda_t^i > 0$ in the presence of attributes $j \neq i$ such that $\lambda_t^j = 0$, the matrix of interactions of i-similar types has all its entries positive. Thus, the analysis is the same as before.

[30] In this case the matrix of interactions is just non-negative. Since it is irreducible, the Perron-

Proof of Proposition 2. Consider the case in which all differences in average initial attitudes are positive. In section 1.4 we comment on the case in which some/all differences associated to attributes $i > 1$ are zero.

To start with, we set the bounds for T_ϵ in expression (1.5). For this purpose recall that $T_\epsilon^{min} = min\{t : \overline{\lambda}_t^1 \geq 1-\epsilon\}$ and $T_\epsilon^{max} = min\{t : \underline{\lambda}_t^1 \geq 1-\epsilon\}$. First, let $\underline{r}_0 = \Delta_0[n]/\Delta_0[1]$. Now, consider $\lambda_t^1 = [1+\sum_{i>1}(r_0^i)^{2^t}]^{-1}$ and replace every $r_t^i = \Delta_t[i]/\Delta_t[1]$ for attributes $i > 1$, with $\underline{r}_0$ to obtain $\overline{\lambda}_t^1 = [1+(n-1)(\underline{r}_0)^{2^t}]^{-1}$. Notice that $\overline{\lambda}_t^1 \geq \lambda_t^1$ at every t. Solving $\overline{\lambda}_t^1 \geq 1-\epsilon$ for t, we get the expression for T_ϵ^{min}, that is, $t = log\left(log\left(\frac{\epsilon}{(1-\epsilon)(n-1)}\right)log(\underline{r}_0)^{-1}\right)\frac{1}{log(2)}$. At every $t' < t$ it follows that $\overline{\lambda}_t^1 \leq 1-\epsilon$, implying that $\lambda_t^1 \leq 1-\epsilon$. Therefore, T_ϵ^{min} is a lower bound for T_ϵ. Second, let $\overline{r}_0 = \Delta_0[2]/\Delta_0[1]$. Replace every r_0^i, for attributes $i > 1$, with $\overline{r}_0$ in λ_t^1. We get $\underline{\lambda}_t^1 = [1+(n-1)(\overline{r}_0)^{2^t}]^{-1}$. Notice that $\underline{\lambda}_t^1 \leq \lambda_t^1$ at every t. Solving $\underline{\lambda}_t^1 \geq 1-\epsilon$ for t, we get the expression for T_ϵ^{max}, that is, $t = log\left(log\left(\frac{\epsilon}{(1-\epsilon)(n-1)}\right)log(\overline{r}_0)^{-1}\right)\frac{1}{log(2)}$. At every $t' > t$ it follows that $\lambda_t^1 \geq \underline{\lambda}_t^1 \geq 1-\epsilon$. Thus, T_ϵ^{max} is an upper bound for T_ϵ. Notice that making T_ϵ^{min} and T_ϵ^{max} positive is always possible, for small enough $\epsilon > 0$.

We now focus on how T_ϵ behaves with respect to changes in the the initial relative salience of attribute 1. Specifically, we do so by proving that λ_t^1 is decreasing in r_0^i. Recall that we consider that the variation in r_0^i comes from varying $\Delta_0[i]$, one at a time. This is done in such a way that $\Delta_0[1] > \Delta_0[2] \geq ... \geq \Delta_0[n] \geq 0$ is preserved in order, as well as in magnitude for differences associated to attributes $j \neq i$. Consider the expression of λ_t^1 above. We have that $\partial\lambda_t^1/\partial r_0^i = -2^t(r_0^i)^{2^t-1}[1+\sum_{i>1}(r_0^i)^{2^t}]^{-2} < 0$. Thus, when r_0^i decreases, at every time t λ_t^1 is higher than before and the time it takes for it to be sufficiently close to its limit is therefore smaller. Being T_ϵ an integer, we thus state that the time it takes for λ_t^1 to be sufficiently close to its limit cannot be higher than before. Finally, notice that λ_t^1 determines the minimum time of convergence for the system as a whole. The reason is that at every time t, λ_t^1 is further away from 1, its limiting value, than any of the remaining homophily values is from 0, its limiting value. To see this notice that at every time t, $\sum_i \lambda_t^i = 1$, then $\lambda_t^1 \geq 1-\epsilon$ implies that $\sum_{i>1}\lambda_t^i \leq \epsilon$. When only λ_t^1 and λ_t^2 are different from zero, then $\lambda_t^1 \geq 1-\epsilon$ implies that $\lambda_t^2 \leq \epsilon$. When λ_t^i is also different from zero for some attributes $i > 2$, then $\lambda_t^1 \geq 1-\epsilon$ implies that $\epsilon/(n-1) \leq \lambda_t^2 < \epsilon$, with $\lambda_t^2 \geq \lambda_t^i$ for any attribute $i > 2$.

■

Proof of footnote 10. We show that $\overline{a}_0 = 0$ implies that $\overline{a}_t = 0$, at every t. By step 5 in the proof of the main Theorem, at every t, W_t is diagonalizable over the same eigenspace. Let G be the projection onto the eigenspace of W_t corresponding to eigen-

Froebenius eigenvalue is equal to the sum of entries of any row in the interaction matrix, which is here always the same. The associated time independent eigenvector is u of size $2^{n-1} \times 1$ with unitary entries.

value 1. Let G^i be the projection onto the eigenspace of W_t corresponding to eigenvalue λ_t^i. By the Spectral Theorem, $W^T a_0 = G a_0 + \sum_{i=1}^n (\prod_{t=0}^T \lambda_t^i) G^i a_0$ (see Meyer (2000), pages 517-520). We proceed by describing how row j of G^i looks like. Denote by G_{jk}^i the jk entry of G^i. It is constructed using eigenvectors in U, in step 5 in the proof of the main Theorem, as follows: $G_{jk}^i = U_{j(i+1)} U'_{(i+1)k}$. In constructing row j of G^i, we fix column $i+1$ in U, i.e., the eigenvector corresponding to λ_t^i, and consider its j entry. Entry j takes value $1/2^{n/2}$ if $i \in A$ and $-1/2^{n/2}$ otherwise. Entry j is multiplied, by the k entries corresponding to row $i+1$ in U', one in a turn. Notice that row $i+1$ of U' is the (transposed) eigenvector associated to λ_t^i. Thus, row j of G^i is just the eigenvector associated to λ_t^i, divided by $1/2^{n/2}$, whenever $i \in A$ and its negative otherwise. Matrix G is constructed in the same way and is composed by ones. Thus, $a_s^A = \overline{a}_0 + 2^{-1} \sum_{i=1}^n (-1)^{1+\mathbf{1}_i} \Delta_0[i] \prod_{t=0}^s \lambda_t^i$, where $\mathbf{1}_i$ is the indicator of type A possessing attribute i. Since there are 2^{n-1} types possessing and lacking every attribute i, respectively, when summing a_s^A for all types, the second term in the previous expression cancels out. Specifically, $\sum_A a_s^A = \sum_A \overline{a}_0 = 2^n \overline{a}_s$. Since $\overline{a}_0 = 0$ then at every time s, $\overline{a}_s = 0$. ■

Chapter 2

Disagreement in Naïve Models of Attitude Formation: Comparative Statics Results

2.1 Introduction

In chapter 1 we proposed a model of attitude formation able to capture the persistence of disagreement. In our approach, individuals, defined by an arbitrary number of attributes, exhibited homophily, the tendency to interact or pay attention to others that are similar to them in terms of shared attributes. We allowed for the possibility that the attention that individuals paid to similar others was co-evolving with attitudes. This co-evolution, that opened the possibility of persistent disagreement, was governed by the salience of attributes. The salience of an attribute was given by the magnitude of the difference in attitudes between the groups of individuals possessing and lacking it. We found that disagreement persisted if and only if there was a unique attribute that became increasingly salient over time.

In our approach individuals had attitudes that were known with certainty, that is, not subject to shocks, and these individuals incorporated others' attitudes following a weighted averaging rule in which relations were symmetric, that is, in which the attention that any pair of individuals paid to each other was the same. In that context, the link between differences in attitudes, determining salience, and attention was given by the Luce form. We endowed the Luce form with the following interpretation: the attention that individuals pay to each other when they share a specific attribute, depends on how big the difference in attitudes associated to this attribute is in relation to the differences associated to the remaining attributes.

In this paper we explore how previous findings react to natural modifications in the assumptions regarding individual behavior. To start with, it might be the case that individuals do not exhibit certain attitudes with respect to a given issue, but that these attitudes are subject to shocks, or are random. In contexts in which individuals aim to learn the true state of the world, randomness might be interpreted as lack of information (noise) regarding the issue at hand, as in Golub and Jackson (2010), as the degree of attitudes' precision, as in DeMarzo et al. (2003), or as experts having subjective probability distributions about the true state, as in DeGroot (1974). In situations in which individuals deal with ideological issues, (maybe) without a well-defined state of the world, we might interpret attitudes' randomness as flexibility or lack of stubbornness. Regarding this point we consider, for the case in which types are defined by two attributes, that initial attitudes of every type are randomly drawn from symmetric continuous distributions. We find that the persistence of disagreement is robust to this type of randomness. In particular, disagreement may now persist across either attribute, being more likely to persist across the one for which the mean of the distribution of the initial difference in attitudes is the highest.

It is also natural to think in situations in which the (intensity of) relations that individuals establish depend(s), not only on the salience of shared attributes, but on the specific nature of these attributes. In fact, McPherson et al. (2001) collect empirical studies documenting how gender homophily is lower when people is young than when old. Gender homophily is also lower for high educated than for low educated people and for Anglos than for African Americans. This might imply, in particular, that pairs of individuals within a relation no longer devote the same amount of attention to each other, that is, that relations are no longer symmetric. We translate these insights in the context of our model as follows: suppose we have four types of individuals, that is, an individual can be either young or old and also either a female or a male. Consider that young people establish less intense relations with the same-gender others than old people. This behavior could emerge in our model when individuals have different sensitivity to differences in attitudes between groups. Specifically, when confronted with information about differences in attitudes between males and females, old people exacerbate the differences in attitudes by gender with respect to young people. Thus, gender is more salient for old people than for young people. As one can observe, the intensity of (gender) relations between pairs of individuals depends on another attribute defining them (youth). We find that disagreement persists across the initially most salient attribute and that its magnitude is higher than in the case in which the attention that any pair of individuals within a relation devote to each other is the same. Also, the time of convergence to the eventual disagreement is lower.

Finally, we explore the more general question of what are the conditions that the evolution of homophily has to satisfy for disagreement to persist. We consider here that individuals are defined by an arbitrary number of attributes. As mentioned, in chapter 1 we used Luce as a particular rule for the evolution of homophily and we discussed how this evolution gave raise to persistent disagreement. That was the case because homophily with respect to the initially most salient attribute increased over time in such a way that the convergence of attitudes to a common value was precluded. In contrast, the constant homophily feature in Golub and Jackson (2012) only affects the speed of convergence to consensus, an outcome that always emerges. We thus find relevant to go beyond in reconciling these two views and in the understanding of what are the homophily patterns that give raise to either persistent disagreement or consensus. We find that in general disagreement persists if and only if the process by which individuals intensify their relations with others with whom they share the initially most salient attribute, is fast enough. More specifically, there are two forces playing a role: on the one hand individuals pay increasing attention to others on the basis of this attribute but on the other hand, they also always pay a positive amount of attention to everyone else. For disagreement to persist it has to be that the first force dominates the second.

The rest of the paper is organized as follows. Section 2.2 discusses random attitudes. Section 2.3 focuses on the case in which individuals' homophilous behavior varies according to a particular attribute defining them. Section 2.4 develops results on disagreement under a general representation of homophily. Section 2.5 concludes. Section 2.6 contains the technical proofs.

2.2 Random attitudes

Let us consider the model in chapter 1 and assume that individuals are composed by two attributes, denoted by 1 and 2. Thus, the type A of an individual is $\{1, 2\}$, $\{1\}$, $\{2\}$ or $\{\emptyset\}$. Let $\tilde{a}_0^A$ be the initial attitude of a type A. Let it follow a symmetric continuous distribution, with mean a_0^A and variance σ_A^2. Initial attitudes of all types are assumed to be independent although not necessarily identically distributed. Let $\tilde{\Delta}_0[1] = 2^{-1}(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} + (\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}))$ and $\tilde{\Delta}_0[2] = 2^{-1}(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} + (\tilde{a}_0^{\{2\}} - \tilde{a}_0^{\{1\}}))$ be the distributions of the initial differences in attitudes associated to attribute 1 and 2, respectively. They have means $\Delta_0[1] = 2^{-1}(a_0^{\{1,2\}} - a_0^{\{\emptyset\}} + (a_0^{\{1\}} - a_0^{\{2\}}))$ and $\Delta_0[2] = 2^{-1}(a_0^{\{1,2\}} - a_0^{\{\emptyset\}} + (a_0^{\{2\}} - a_0^{\{1\}}))$, respectively, and the same variance, $\sum_A \sigma_A^2/4$. We assume without loss of generality that these means are such that $\Delta_0[1] \geq \Delta_0[2] \geq 0$. In linking homophily and salience, we discuss the Luce form, as in chapter 1. Thus:

$$\tilde{\lambda}_t^1 = \frac{|\tilde{\Delta}_t[1]|}{|\tilde{\Delta}_t[1]| + |\tilde{\Delta}_t[2]|} \text{ and } \tilde{\lambda}_t^2 = \frac{|\tilde{\Delta}_t[2]|}{|\tilde{\Delta}_t[1]| + |\tilde{\Delta}_t[2]|}.$$

Notice that the homophily values could, in principle, be positive or negative, depending on the realization of the random variables $\tilde{\Delta}_t[1]$ and $\tilde{\Delta}_t[2]$. As we assume that the only aspect that matters is the magnitude of the differences in attitudes and not their sign, we work with differences in absolute value.

As a preview of the results, we find that the persistence of disagreement is robust to this type of randomness. In contrast with the deterministic case, in general disagreement persists across either attribute with positive probability. Disagreement will persist across attribute 1 with probability equal to one when the minimum among all possible realizations of $|\tilde{\Delta}_0[1]|$ is higher than the maximum among all possible realizations of $|\tilde{\Delta}_0[2]|$. We now discuss how the likelihood that disagreement persists across either attribute depends on the features of the distributions of initial differences in attitudes associated to these attributes. The results are as follows:

Proposition 1. *In general disagreement persists across either attribute* 1 *or* 2 *with positive probability. Also, the difference between the probability that disagreement persists across attribute 1 and the one that it does across attribute 2 is given by:*

$$(2P(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0) - 1)(P(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0) - P(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} < 0)). \tag{2.1}$$

Specifically, disagreement across attribute 1 *is at least as likely as disagreement across attribute* 2*. Both events are equally likely if and only if both attributes are, initially, equally salient in mean (that is, if and only if* $\Delta_0[1] = \Delta_0[2]$*, or equivalently,* $a_0^{\{1\}} - a_0^{\{2\}} = 0$*) whereas disagreement across attribute* 1 *is the most likely event if and only if attribute* 1 *is, initially, the most salient in mean (that is, if and only if,* $\Delta_0[1] > \Delta_0[2]$*, or equivalently,* $a_0^{\{1\}} - a_0^{\{2\}} > 0$*).*

Notice that expression (2.1) is non-negative. That is so because the (symmetric) distributions of $\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}}$ and $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}$ have non-negative means.[1] Furthermore, regardless of the variance of $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}$ the probability that it takes values higher or equal than zero, increases with its mean. Thus the likelihood of disagreement across attribute 1 increases.[2]

Notice that since we focus on perturbations in initial attitudes, once they are realized, eventual attitudes acquire the same form as the ones in the main Theorem in chapter 1. We now clarify this point. For this purpose let $\tilde{a}_\infty^A$ denote the eventual attitude of a type A:

Remark 1. *Suppose that disagreement persists across attribute* 1*. Then,* $\tilde{a}_\infty^A = \overline{a}_0 + 2^{-1}\left(1 - |\tilde{\Delta}_0[2]|/|\tilde{\Delta}_0[1]|\right)\tilde{\Delta}_0[1]$ *if* $1 \in A$ *and* $\tilde{a}_\infty^A = \overline{a}_0 - 2^{-1}\left(1 - |\tilde{\Delta}_0[2]|/|\tilde{\Delta}_0[1]|\right)\tilde{\Delta}_0[1]$ *if* $1 \notin A$*.*

Remark 2. *Suppose that disagreement persists across attribute* 2*. Then,* $\tilde{a}_\infty^A = \overline{a}_0 + 2^{-1}\left(1 - |\tilde{\Delta}_0[1]|/|\tilde{\Delta}_0[2]|\right)\tilde{\Delta}_0[2]$ *if* $2 \in A$ *and* $\tilde{a}_\infty^A = \overline{a}_0 - 2^{-1}\left(1 - |\tilde{\Delta}_0[1]|/|\tilde{\Delta}_0[2]|\right)\tilde{\Delta}_0[2]$ *if* $2 \notin A$*.*

[1] See the proof of Proposition 1.

[2] When expression (2.1) is strictly positive an increase in the mean of $\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}}$ also increases its value.

The following examples illustrate these findings and related aspects. In example 1 we pin down the probability that disagreement persists across either attribute when initial attitudes are uniformly distributed. In example 2, initial attitudes are normally distributed. We document how the probability that disagreement persists across attribute 1 may increase or decrease with the mean and the variance of initial attitudes:

Example 1. First, let initial attitudes be such that $a_0^{\{1,2\}} \sim U[0,1]$, $\tilde{a}_0^{\{1\}} \sim U[-1,1]$, $\tilde{a}_0^{\{2\}} \sim U[-1,1]$ and $\tilde{a}_0^{\{\emptyset\}} \sim U[-1,1]$. Thus, $\tilde{\Delta}_0[1]$ and $\tilde{\Delta}_0[2]$ have means $\Delta_0[1] = \Delta_0[2] = 0.25$. From Proposition 1, disagreement across either attribute is equally likely. To see this recall that the probability that disagreement persists across attribute 1 minus the probability that it does across attribute 2 depends on $\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}}$ and $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}$. As $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}$ follows a (symmetric) triangular distribution with mean zero, this difference in probabilities is zero. Second, let initial attitudes be such that $a_0^{\{1,2\}} \sim U[0,1]$, $\tilde{a}_0^{\{1\}} \sim U[0,1]$, $\tilde{a}_0^{\{2\}} \sim U[-1,1]$ and $\tilde{a}_0^{\{\emptyset\}} \sim U[-1,1]$. Thus, $\tilde{\Delta}_0[1]$ and $\tilde{\Delta}_0[2]$ have means $\Delta_0[1] = 0.5$ and $\Delta_0[2] = 0$, respectively. From Proposition 1, disagreement across attribute 1 is the most likely event. As above we focus on the distributions of $\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}}$ and $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}$. Let $y \equiv \tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}}$. It follows a triangular distribution with density:

$$f(y) = \begin{cases} \dfrac{1+y}{2} & if \quad -1 < y < 0 \\ 0.5 & if \quad 0 \leq y \leq 1 \\ 1 - \dfrac{y}{2} & if \quad 1 < x < 2 \end{cases} .$$

Thus, $P(y \geq 0) = 1 - \int\limits_{y=-1}^{y=0} \dfrac{1+y}{2} = 0.75$. Also, let $z \equiv \tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0$. Notice that it follows the same distribution as y. Thus, $P(z \geq 0) = 0.75$ as well. In this case expression (2.1) equals 0.25. Thus disagreement persists across attribute 1 and 2 with probabilities 0.625 and 0.375, respectively.

Example 2. Let initial attitudes be normally distributed with means such that $\Delta_0[1] \geq \Delta_0[2] > 0$ and variances equal to one.[3] In the first figure we depict the probability that disagreement persist across attribute 1, as a function of the mean of the distribution of difference in attitudes associated it. In particular we keep $\Delta_0[2]$ and increase $\Delta_0[1]$. On the x-axis we depict the ratio $x = \Delta_0[1]/\Delta_0[2]$ and on the y-axis, the probability of disagreement across attribute 1. We observe a positive relation. In the second figure we depict the distribution of eventual attitudes when disagreement persists across attribute 1 for the case in which $\Delta_0[1]$ =6 and $\Delta_0[2] = 2$ and thus $x = 3$.[4]

[3] As stated in chapter 1 the results remain the same if we consider that initial attitudes are defined over the entire real line.

[4] We simulate the process 1500 times for any configuration of the ratio x. The Matlab code is available upon request.

Figure 1. Probability that disagreement persists across attribute 1

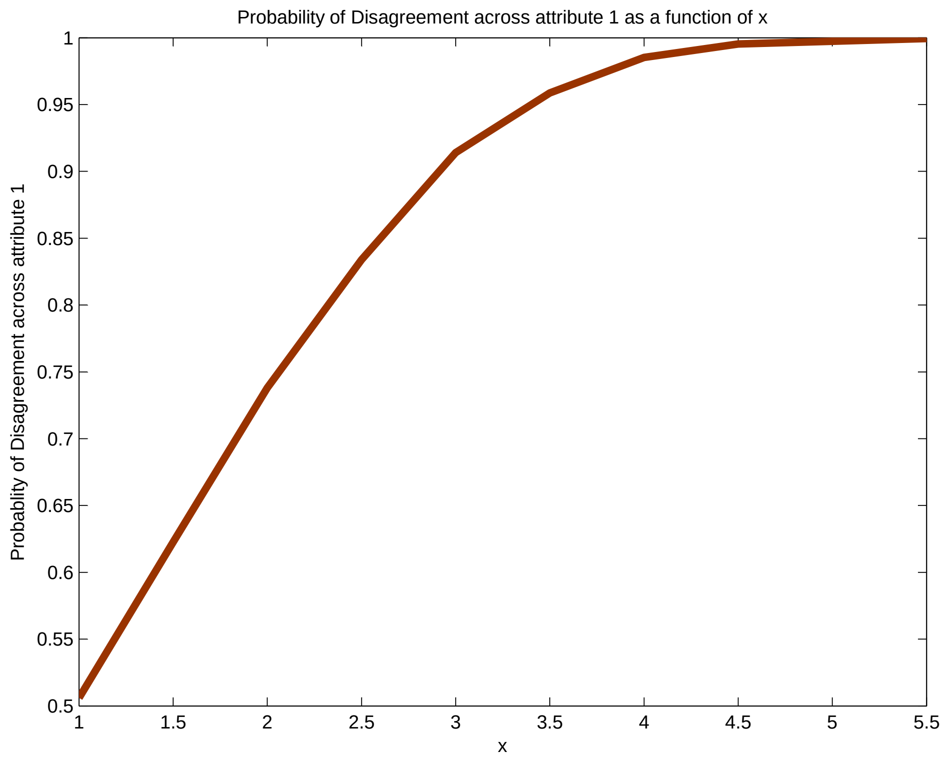
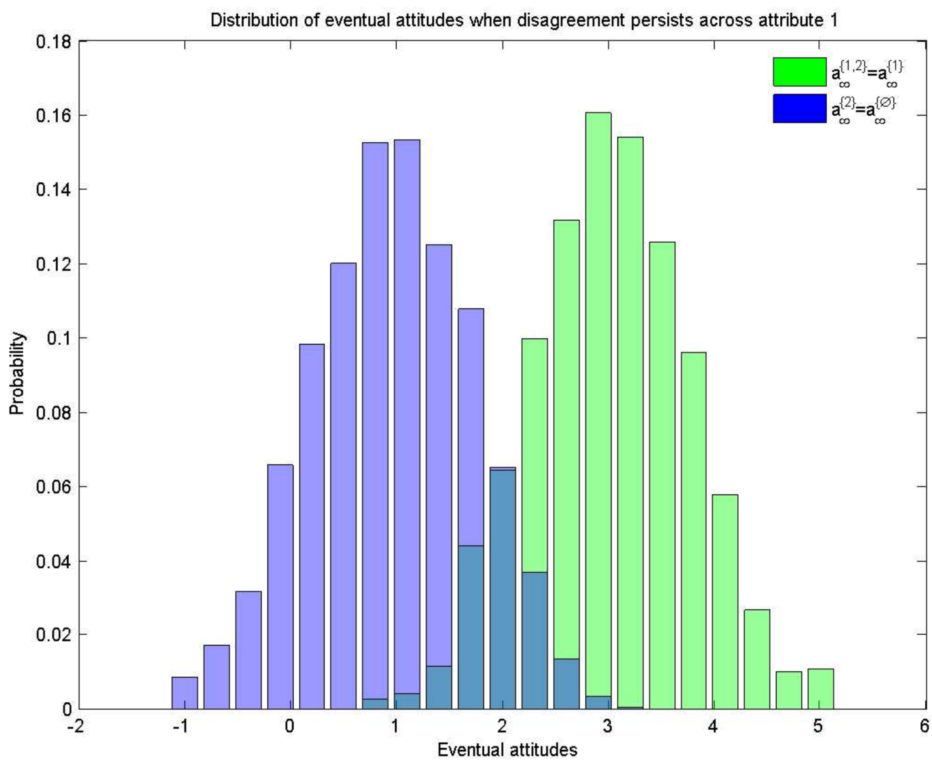


Figure 2. Distribution of eventual attitudes



Within this example, it is also worth illustrating how the variances of the distributions of initial attitudes may play a role in determining the likelihood of disagreement. Let us consider that initial attitudes are normally distributed with the same means as above, for the cases in which $\Delta_0[1] > \Delta_0[2] \geq 0$. In contrast, let the variances of these random variables, instead of being all equal to one, be such that $\tilde{a}_0^{\{1\}'} - \tilde{a}_0^{\{2\}'}$ and/or $\tilde{a}_0^{\{1,2\}'} - \tilde{a}_0^{\{\emptyset\}'}$ are mean preserving spreads of $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}$ and/or $\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}}$, respectively. Thus, $0.5 < P(\tilde{a}_0^{\{1\}'} - \tilde{a}_0^{\{2\}'} \geq 0) < P(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0)$ and/or $0.5 < P(\tilde{a}_0^{\{1,2\}'} - \tilde{a}_0^{\{\emptyset\}'} \geq 0) < P(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0)$.[5] Notice that expression (2.1) in Proposition 1 has now lower value than before, meaning that disagreements across either attribute are closer to being equally likely.

To conclude, disagreement manifests as two groups holding different eventual attitudes. In remark 1 (respectively remark 2), the difference in average eventual attitudes associated to attribute 1 (respectively 2) persists whereas the one associated to attribute 2 (respectively 1) is zero.[6]

2.3 Non-symmetric Homophily

Let us consider the model in chapter 1 and assume that individuals are composed by two attributes. Let attribute 1 refer to gender (for instance, possessing it means being female whereas lacking it means being male) and attribute 2 refer to youth (for instance, possessing it means being young whereas lacking it means being old). Recall that at every time t, $\lambda_t^1 = \dfrac{\Delta_t[1]}{\Delta_t[1] + \Delta_t[2]}$ and that $\lambda_t^2 = \dfrac{\Delta_t[2]}{\Delta_t[1] + \Delta_t[2]}$. Recall also that, without loss of generality we consider that differences in average initial attitudes are such that, $\Delta_0[1] \geq \Delta_0[2] \geq 0$.

Suppose that old types (those lacking attribute 2) are more homophilous with respect to gender than young types (those possessing attribute 2). In our context this means that old types are more sensitive to differences in attitudes associated to gender

[5]Notice that $\Delta_0[1] = \Delta_0[2] \geq 0$ holds when $a_0^{\{1\}} - a_0^{\{2\}} = 0$. In this case, regardless of the variances, disagreement across either attribute is equally likely. See the proof of Proposition 1.

[6]When disagreement persists across attribute 1, the difference in average eventual attitudes associated to attribute 1 is $\tilde{\Delta}_\infty[1] = (2^{n-1})^{-1}(\sum_{A:i\in A}\tilde{a}_\infty^A - \sum_{A:1\notin A}\tilde{a}_\infty^A) = (2^{n-1})^{-1}2^{n-1}(\tilde{a}_\infty^A : 1 \in A - \tilde{a}_\infty^A : 1 \notin A) = |\tilde{\Delta}_0[1]| - |\tilde{\Delta}_0[2]|$ (respectively $|\tilde{\Delta}_0[2]| - |\tilde{\Delta}_0[1]|$) when $\tilde{\Delta}_0[1] \geq 0$ (respectively $\tilde{\Delta}_0[1] < 0$). Since disagreement across attribute 1 persists when $|\tilde{\Delta}_0[1]| > |\tilde{\Delta}_0[2]|$, $\tilde{\Delta}_\infty[1]$ has either positive or negative support. Furthermore, the distribution of the difference in average eventual attitudes associated to attribute 2 is degenerated at zero. To see this notice that within the 2^{n-1} types possessing attribute 1 there are 2^{n-2} types possessing and lacking attribute 2, respectively. The same happens within the 2^{n-1} types lacking attribute 1, hence, $\tilde{\Delta}_\infty[2] = (2^{n-1})^{-1}(\sum_{A:2\in A}\tilde{a}_\infty^A - \sum_{A:2\notin A}\tilde{a}_\infty^A) = 2^{-1}(2^{n-2})^{-1}2^{n-2}(\tilde{a}_\infty^A : 1 \in A, 2 \in A + \tilde{a}_\infty^A : 1 \notin A, 2 \in A) - \tilde{a}_\infty^A : 1 \in A, 2 \notin A + \tilde{a}_\infty^A : 1 \notin A, 2 \notin A) = 2^{-1}2(\overline{a}_0 - \overline{a}_0) = 0$. The analysis is the same when disagreement persists across attribute 2.

than young types. Notice that given our structure of attention, as old types are more homophilous with respect to gender than young types, they are also less homophilous with respect to youth than young types. We now formally describe the matrix of interactions as:

$$\mathrm{W}_t = \frac{1}{2}\begin{array}{c} \begin{array}{cccc} \{1,2\} & \{1\} & \{2\} & \{\emptyset\} \end{array} \\ \begin{bmatrix} \lambda_t^1+\lambda_t^2 & \lambda_t^1 & \lambda_t^2 & 0 \\ \beta_t^1 & \beta_t^1+\beta_t^2 & 0 & \beta_t^2 \\ \lambda_t^2 & 0 & \lambda_t^1+\lambda_t^2 & \lambda_t^1 \\ 0 & \beta_t^2 & \beta_t^1 & \beta_t^1+\beta_t^2 \end{bmatrix} \end{array} \begin{array}{c} \\ \{1,2\} \\ \{1\} \\ \{2\} \\ \{\emptyset\} \end{array},$$

where at every time t, β_t^1 and β_t^2 are assumed to depend on $\Delta_t[1]$ and $\Delta_t[2]$ and be such that old people exacerbate homophily towards gender. Specifically, let β_t^1, $\beta_t^2 \in [0,1]$ be such that $\beta_t^1 + \beta_t^2 = 1$ and $\beta_t^1 > \lambda_t^1$, $\beta_t^2 < \lambda_t^2$. In particular, $\beta_t^1 = 1$ whenever $\lambda_t^1 = 1$ and thus $\beta_t^2 = 0$ whenever $\lambda_t^2 = 0$. For the ease of exposition one can consider that β_t^1 and β_t^2 are transformations of λ_t^1 and λ_t^2, respectively, such that old people are more sensitive to the salience of gender than young people. For any specific homophily structure defined in these terms it follows that interactions, represented by the matrix above, become non-symmetric. Finally, the law of motion of attitudes is given by $a_{t+1} = W_t a_t$, as in chapter 1.

We now elaborate on the consequences that this new type of interactions has on disagreement. The results are as follows:

Proposition 2. *For every configuration of initial attitudes, eventual ones always exist and exhibit disagreement across attribute 1. Specifically, types hold the same eventual attitude if and only if they share attribute* 1*. Furthermore, eventual disagreement is larger than in the symmetric case where, at every* t, $\beta_t^1 = \lambda_t^1$ *and* $\beta_t^2 = \lambda_t^2$.

The extreme version of this new scenario is such that at every time t, and regardless of the magnitude of differences in attitudes, $\beta_t^1 = 1$ and thus, $\beta_t^2 = 0$. Then attribute 2 (youth) does not play any role for old types, namely, $\{1\}$ and $\{\emptyset\}$, and they only pay attention to others, based on the gender dimension, that is, they only pay attention to types $\{1,2\}$ and $\{2\}$, respectively (and to themselves).

In general this process will give raise to disagreement across attribute 1 even in the case in which the differences in average initial attitudes are the same, that is, when $\Delta_0[1] = \Delta_0[2]$. That is so because attribute 1 deserves even more attention now that in the case in which relations are symmetric. The process also converges faster than before to the eventual disagreement. The following example illustrates these findings:

Example 3. Consider that initial attitudes are $a_0' = [0.8 \;\; 0.2 \;\; -0.05 \;\; -0.95]$. In this case $\Delta_0[1] = 1$ and $\Delta_0[2] = 0.75$, and thus, as stated in the main Theorem in chapter 1, disagreement persists and manifests in two groups defined according to the possession or lack of attribute 1, holding different eventual attitudes. These are, $a'_\infty = [0.125 \quad 0.125 \quad -0.125 \quad -0.125]$. Let $\alpha > 1$ and set $\beta_t^2 = (\lambda_t^2)^\alpha < \lambda_t^2$, thus $\beta_t^1 = 1 - (\lambda_t^2)^\alpha > 1 - \lambda_t^2 = \lambda_t^1$. In particular, for $\alpha = 2$ we have that $a'_\infty = [0.341 \ \ 0.341 \ -0.226 \ -0.226]$. The eventual disagreement, measured as the difference in average eventual attitudes between the groups of individuals possessing and lacking attribute 1, is 0.25 in the symmetric case and 0.567 in the case described here. The later would also increase with the value of α.

2.4 A general representation of homophily

In this section we relate the persistence of disagreement to the properties and evolution of homophily. For this purpose we use the model in chapter 1 but we define homophily values in broader terms. Specifically, let γ_t^i be the homophily value associated to attribute i at time t. Let this value depend on the differences in average attitudes associated to (possibly) all attributes. As in chapter 1 we assume that at every time t, γ_t^i is non-negative and we normalize to one the total amount of attention that every individual devotes to others. It then has to be the case that at every time t the sum of these homophily values is one, formally, $\sum_i \gamma_t^i = 1$.[7] We finally assume that the homophily values satisfy two properties that deal with the monotonicity aspects of attention with respect to the differences in attitudes. The first one states that the attention that every attribute enjoys is positive if and only if the difference in attitudes across it, is positive. The second one states that if attribute i exhibits a higher difference in attitudes than attribute j then the former enjoys higher attention than the latter:

Within differences monotonicity (WDM). $\Delta_t[i] = 0$ implies that $\gamma_t^i = 0$ and $\Delta_t[i] > 0$ implies that $\gamma_t^i > 0$.[8]

Across differences monotonicity (ADM). $\Delta_t[1] \geq \Delta_t[2] \geq \cdots \geq \Delta_t[n] \geq 0$ implies that $\gamma_t^1 \geq \gamma_t^2 \geq \cdots \geq \gamma_t^n \geq 0$.

We also set the technical condition that $\lim_{t\longrightarrow\infty} \gamma_t^i$ exists for every attribute i and that $\lim_{t\longrightarrow\infty} \sum_i \gamma_t^i = \sum_i \lim_{t\longrightarrow\infty} \gamma_t^i = 1$.

We now state the condition for the persistence of disagreement and provide its form:

Proposition 3. *For every configuration of initial attitudes, eventual ones always exist. They exhibit disagreement if and only if homophily based on attribute* 1, *approaches*

[7] See chapter 1, section 1.2.

[8] When $\Delta_t[i] = 0$ for every attribute i, we set $\gamma_t^i = 1/n$.

value 1 *sufficiently fast (that is, if and only if* $\sum_{t=0}^{\infty} log\gamma_t^1$ *exists). In this case, eventual attitudes are such that, for every type A:*

$$|a_\infty^A| = \frac{1}{2}\tau(W^\infty)\Delta_0[1]$$

where $\tau(W^\infty) \in (0,1]$. *Furthermore,* $a_\infty^A > 0$ *if and only if* $1 \in A$.

In general disagreement persists whenever the process by which individuals progressively intensify their relations with others similar to them in attribute 1 is fast enough. Intuitively there are two forces playing a role: on the one hand individuals pay increasing attention to others on the basis of attribute 1 but on the other hand, they also pay a positive amount of (possibly indirect) attention to everyone else. For disagreement to persist, it has to be that the first force dominates the second. Needless to say that the Luce form, presented in chapter 1, satisfies these requirements. As an illustration, for the case with two attributes $\sum_{t=0}^{\infty} log\gamma_t^1 = log(1 - \Delta_0[2]/\Delta_0[1])$ and $\tau(W^\infty) = 1 - \Delta_0[2]/\Delta_0[1]$, with $\Delta_0[1] > \Delta_0[2] \geq 0$.

Disagreement materializes in two groups of thinking, defined according to whether individuals possess or lack attribute 1. We cannot specify the closed form expression for the ergodicity coefficient $\tau(W^\infty)$ in this case, since it depends on the particular functional form for the homophily values. We just set $\tau(W^\infty) = \lim_{T\to\infty}\prod_{t=0}^{T}\gamma_t^1$ in this case.

The following examples illustrate the requirement in the proposition above. For this purpose, we consider updating rules that are mainly based on modifications of the Luce form in expression (1.2), with the exception of example 7. Example 4 deals with a scenario in which consensus is achieved whereas in examples 5 to 7 disagreement persists.

Example 4. Eventual consensus. Consider the following updating rule:

$$\gamma_t^1 = \begin{cases} \dfrac{\Delta_t[1]}{\Delta_t[1] + \Delta_t[2]} & if \quad \gamma_{t-1}^1 < H \in [0,1) \\ \gamma_{t-1}^1 & if \quad \gamma_{t-1}^1 \geq H \in [0,1) \end{cases}.$$

Let $\gamma_t^2 = 1 - \gamma_t^1$ at every time t. Under this rule individuals use Luce to determine the attention they pay to others, but whenever a level H of homophily has been reached, they are no longer sensitive to changes in differences in attitudes. In this case interactions become static from some point in time on, and thus, individuals do not become homophilous exclusively with respect to attribute 1. The requirements in the proposition above are therefore not satisfied and consensus will eventually emerge.

Example 5. The persistence of disagreement. Let initial attitudes be $a_0' = [0.8 \;\; 0.2 \;\; -0.05 \;\; -0.95]$. Thus, the difference in average initial attitudes associated to attribute 1 is $\Delta_0[1] = 0.5(0.8 + 0.2) - 0.5(-0.05 - 0.95) = 1$ and the one associated to attribute 2 is $\Delta_0[2] = 0.5(0.8 - 0.05) - 0.5(0.2 - 0.95) = 0.75$. Consider the generalized Luce form, $\gamma_t^1 = \dfrac{\Delta_t[1]^\delta}{\Delta_t[1]^\delta + \Delta_t[2]^\delta}$ and $\gamma_t^2 = \dfrac{\Delta_t[2]^\delta}{\Delta_t[1]^\delta + \Delta_t[2]^\delta}$. When $\delta = 1.2$, $\gamma_0^1 = 0.58$ and $\gamma_0^2 = 0.42$. The entries in the interaction matrices evolve as follows:

$$W_0 = \begin{bmatrix} 0.5 & 0.3 & 0.2 & 0 \\ 0.3 & 0.5 & 0 & 0.2 \\ 0.3 & 0 & 0.5 & 0.3 \\ 0 & 0.2 & 0.3 & 0.5 \end{bmatrix},\ W_1 = \begin{bmatrix} 0.5 & 0.34 & 0.16 & 0 \\ 0.34 & 0.5 & 0 & 0.16 \\ 0.16 & 0 & 0.5 & 0.34 \\ 0 & 0.16 & 0.34 & 0.5 \end{bmatrix}, ..., \lim_{t\to\infty} W_t = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}.$$

$$\text{Also, } W^\infty = \begin{bmatrix} 0.33 & 0.33 & 0.17 & 0.17 \\ 0.33 & 0.33 & 0.17 & 0.17 \\ 0.17 & 0.17 & 0.33 & 0.33 \\ 0.17 & 0.17 & 0.33 & 0.33 \end{bmatrix} \text{ and } W^\infty \text{ times } a_0 = \begin{bmatrix} 0.8 \\ 0.2 \\ -0.05 \\ -0.95 \end{bmatrix} \text{ is } a_\infty = \begin{bmatrix} 0.16 \\ 0.16 \\ -0.16 \\ -0.16 \end{bmatrix}.$$

In this case $\tau(W^\infty) = 0.32$.[9] Notice that the relation between attributes 1 and 2, summarized in $(\Delta_0[2]/\Delta_0[1])^{1.2}$, exacerbates with respect to the case in which $\delta = 1$.

Example 6. The persistence of disagreement. Consider the case in which $\gamma_t^1 = \dfrac{\beta\Delta_t[1]}{\beta\Delta_t[1] + \delta\Delta_t[2]}$ and $\gamma_t^2 = \dfrac{\delta\Delta_t[2]}{\beta\Delta_t[1] + \delta\Delta_t[2]}$, with $\beta > \delta > 0$. Notice that the relation between attributes 1 and 2, that is, $\delta\Delta_0[2]/\beta\Delta_0[1]$, exacerbates with respect to the case in which $\beta = \delta$. This process leads to disagreement for any configuration of initial attitudes, that is, even in the case in which $\Delta_0[1] = \Delta_0[2]$. The reason is that $\gamma_0^1 > \gamma_0^2$ and $\Delta_1[1] = \gamma_0^1\Delta_0[1] > \Delta_1[2] = \gamma_0^2\Delta_0[2]$, thus form $t = 1$ the main Theorem in chapter 1, and thus the requirements in the proposition above, apply.

Example 7. The persistence of disagreement. Let initial attitudes be $a_0' = [0.8 \quad 0.2 \quad -0.05 \quad -0.95]$. Thus, the difference in average initial attitudes associated to attribute 1 is $\Delta_0[1] = 1$ and the one associated to attribute 2 is $\Delta_0[2] = 0.75$, as in example 5. Consider the following updating rule:

$$\gamma_t^1 = \begin{cases} 0.5 & if \quad \Delta_t[1] = \Delta_t[2] \geq 0 \\ 1 & if \quad \Delta_t[1] > \Delta_t[2] = 0 \\ 0 & if \quad \Delta_t[2] > \Delta_t[1] = 0 \\ \left(\dfrac{\Delta_t[1]}{\Delta_t[2]}\right)^\alpha & if \quad \Delta_t[2] > \Delta_t[1] > 0 \\ 1 - \left(\dfrac{\Delta_t[2]}{\Delta_t[1]}\right)^\beta & if \quad \Delta_t[1] > \Delta_t[2] > 0 \end{cases}.$$

[9]The entries of W^∞ are a function of $\tau(W^\infty)$, thus we can recover its value once W^∞ is known.

We set α and β such that $(\Delta_t[1]/\Delta_t[2])^\alpha$ and $(\Delta_t[2]/\Delta_t[1])^\beta$ are smaller than one half. Let $\gamma_t^2 = 1 - \gamma_t^1$, at every time t. Let us set, for instance, $\beta = 2.5$.[10] The entries in the interaction matrix evolve as follows:

$$W_0 = \begin{bmatrix} 0.5 & 0.26 & 0.24 & 0 \\ 0.26 & 0.5 & 0 & 0.24 \\ 0.24 & 0 & 0.5 & 0.26 \\ 0 & 0.24 & 0.26 & 0.5 \end{bmatrix}, \; W_1 = \begin{bmatrix} 0.5 & 0.28 & 0.22 & 0 \\ 0.28 & 0.5 & 0 & 0.22 \\ 0.22 & 0 & 0.5 & 0.28 \\ 0 & 0.22 & 0.28 & 0.5 \end{bmatrix}, \ldots, \lim_{t\to\infty} W_t = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}.$$

$$\text{Also, } W^\infty = \begin{bmatrix} 0.31 & 0.31 & 0.19 & 0.19 \\ 0.31 & 0.31 & 0.19 & 0.19 \\ 0.19 & 0.19 & 0.31 & 0.31 \\ 0.19 & 0.19 & 0.31 & 0.31 \end{bmatrix} \text{ and } W^\infty \text{ times } a_0 = \begin{bmatrix} 0.8 \\ 0.2 \\ -0.05 \\ -0.95 \end{bmatrix} \text{ is } a_\infty = \begin{bmatrix} 0.12 \\ 0.12 \\ -0.12 \\ -0.12 \end{bmatrix}.$$

In this case $\tau(W^\infty) = 0.24$.

2.5 Conclusions

We have explored the model of attitude formation presented in chapter 1 allowing for natural modifications in the initial assumptions about individuals' behavior. The aim has been to identify how these natural changes affect previous findings regarding the persistence of disagreement.

With respect to non-symmetric homophily, it would be interesting to explore the case in which, on the one hand attribute 1 is, initially, at least equally salient as attribute 2, that is, $\Delta_0[1] \geq \Delta_0[2]$, but on the other hand some individuals pay more attention than before to others similar to them in attribute 2, that is, at every time t $\beta_t^2 > \lambda_t^2$ and thus, $\beta_t^1 < \lambda_t^1$. In particular, $\beta_t^2 = 1$ whenever $\lambda_t^2 = 1$ and thus $\beta_t^1 = 0$ whenever $\lambda_t^1 = 0$. Our conjecture is that in this case two situations might arise. First, it might be that differences associated to attribute 1 remain the highest in subsequent points in time. In this case the challenge is to carefully study whether the process leads to eventual consensus or persistent disagreement. Second, it might also be the case that differences in attitudes associated to attribute 2 become the highest at some point in time. In this case disagreement persists across this attribute.

Related to the last scenario consider, as a first illustration, the particular case in which $\Delta_0[1] = \Delta_0[2] > 0$ but $\beta_t^1 < \lambda_t^1$. Then it directly follows that $\Delta_1[1] < \Delta_1[2]$. Thus, $\lambda_1^1 < 2^{-1} < \lambda_1^2$ and $\beta_1^1 < 2^{-1} < \beta_1^2$. The analysis from $t = 1$ on is the same

[10] Since at every time t it is the case that $\Delta_t[1] > \Delta_t[2] > 0$, we do not specify any value for α. Also, since $\Delta_t[1]/\Delta_t[2]$ (respectively $\Delta_t[2]/\Delta_t[1]$) is decreasing over time whenever $\Delta_t[2] > \Delta_t[1]$ (respectively $\Delta_t[2] < \Delta_t[1]$) we set α and β to be constant. See steps 1 and 7 in the proof of the main Theorem in chapter 1.

than the one carried out in building Proposition 2, now applied to differences across attribute 2. Thus, disagreement persists across this attribute. As a second illustration, consider the particular case in which $\Delta_0[1] > \Delta_0[2] > 0$, that is, attribute 1 is initially the most salient, but at every time t $\beta_t^1 = 0$ and thus $\beta_t^2 = 1$. In this case disagreement also persists across attribute 2.[11]

Another possibility is to study the case in which, some types exhibit constant homophily, that is, the attention they pay to similar others does not co-evolve with attitudes. This is the case if, for instance, $\beta_t^1 = \alpha$ in (0,1), at every time t. In this case, when the limiting matrix of interactions, namely, $\lim_{t\to\infty} W_t$, exists with $\lim_{t\to\infty} \lambda_t^1 \in [0,1)$, this matrix would be such that no pair of rows are orthogonal.[12] Following Leizarowitz (1992), consensus will eventually emerge in this case. When $\lim_{t\to\infty} \lambda_t^1 \in (0,1)$, eventual interactions allow attitudes to flow from every individual to any other. In other words, the graph describing these interactions is strongly connected.

2.6 Appendix. Proofs

Proof of Proposition 1. We compute here the probability that disagreement persists across attribute 1 and across attribute 2. For this purpose, let us first focus on the case in which $\Delta_0[1] > \Delta_0[2] \geq 0$, or equivalently, $a_0^{\{1,2\}} - a_0^{\{\emptyset\}} + a_0^{\{1\}} - a_0^{\{2\}} > a_0^{\{1,2\}} - a_0^{\{\emptyset\}} + a_0^{\{2\}} - a_0^{\{1\}}$. Notice that $\Delta_0[1] - \Delta_0[2] = a_0^{\{1\}} - a_0^{\{2\}}$. Thus, $a_0^{\{1\}} - a_0^{\{2\}} > 0$ has to hold. Since $\Delta_0[2] \geq 0$ and $a_0^{\{2\}} - a_0^{\{1\}} < 0$, then $a_0^{\{1,2\}} - a_0^{\{\emptyset\}} \geq a_0^{\{1\}} - a_0^{\{2\}} > 0$ has to hold as well.

Now, consensus emerges whenever $|\tilde{\Delta}_0[1]| = |\tilde{\Delta}_0[2]|$ and disagreement persists across attribute 1 (respectively attribute 2) whenever $|\tilde{\Delta}_0[1]| > |\tilde{\Delta}_0[2]|$ (respectively $|\tilde{\Delta}_0[1]| < |\tilde{\Delta}_0[2]|$). To see this notice that once initial attitudes are realized, the process exactly mimics the one presented in chapter 1. In what follows we describe the probability that either consensus emerges or disagreement persists. The probability that $|\tilde{\Delta}_0[1]| = |\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} + \tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}| = |\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} + \tilde{a}_0^{\{2\}} - \tilde{a}_0^{\{1\}}| = |\tilde{\Delta}_0[2]|$ is zero. That is so because this expression holds when exactly $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} = 0$ and/or $\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} = 0$. Since these differences follow continuous distributions, the proba-

[11] From the proof of Proposition 2, we have that $\Delta_t[1] = (\lambda_{t-1}^1/2)\Delta_{t-1}[1]$ and $\Delta_t[2] = ((\lambda_{t-1}^2 + 1)/2)\Delta_{t-1}[2]$. Thus, $\lambda_1^1 = \dfrac{\Delta_1[1]}{\Delta_1[1] + \Delta_1[2]} = \dfrac{(\lambda_{t-1}^1/2)\Delta_0[1]}{(\lambda_{t-1}^1/2)\Delta_0[1] + (1 - \lambda_{t-1}^1/2)\Delta_0[2]}$. This is equivalent to $\lambda_1^1 = (1 + r_0^2(2 - \lambda_0^1)/\lambda_0^1)^{-1}$ where $r_0^2 = \Delta_0[2]/\Delta_0[1]$. Notice that $\lambda_0^1 = (1 + r_0^2)^{-1}$. Since $\lambda_0^1 < 1$, thus $\lambda_1^1 < \lambda_0^1$. At $t = 2$, $\lambda_2^1 = \dfrac{1}{1 + r_0^2((2 - \lambda_1^1)/\lambda_1^1)(2 - \lambda_0^1)/\lambda_0^1}$. Since $\lambda_1^1 < \lambda_0^1$ it follows that $\lambda_2^1 < \lambda_1^1$. In general, at every point it time, $\lambda_{t+1}^1 < \lambda_t^1$, so that λ_t^1 tends to 0 and λ_t^2 tends to 1. Thus, for sufficiently large t, it has to be that $\lambda_t^2 > \lambda_t^1$ and thus $\Delta_t[2] > \Delta_t[1]$. From this point on, we apply to proof the proof of Proposition 2 to attribute 2.

[12] Notice that for $\lim_{t\to\infty} W_t$ to exist it remains to be checked that $\lim_{t\to\infty} \lambda_t^1$ exists.

bility that this happens is zero.[13] Disagreement persists across attribute 1 whenever $|\tilde{\Delta}_0[1]| = |\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} + \tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}| > |\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} + \tilde{a}_0^{\{2\}} - \tilde{a}_0^{\{1\}}| = |\tilde{\Delta}_0[2]|$. This expression is satisfied when $\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0$ and $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0$, or $\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} < 0$ and $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} < 0$ hold. Thus, $P(|\tilde{\Delta}_t[1]| > |\tilde{\Delta}_t[2]|) = P(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0 \cap \tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0) + P(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} < 0 \cap \tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} < 0)$. Since $\tilde{a}_0^A$ are independent to each other, this is equivalent to:

$$P(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0)P(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0) + P(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} < 0)P(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} < 0).$$

On the contrary, disagreement persists across attribute 2 whenever $|\tilde{\Delta}_0[1]| = |\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} + \tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}| < |\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} + \tilde{a}_0^{\{2\}} - \tilde{a}_0^{\{1\}}| = |\tilde{\Delta}_0[2]|$. This expression is satisfied when $\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} < 0$ and $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0$, or $\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0$ and $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} < 0$ hold. Then $P(|\tilde{\Delta}_0[1]| < |\tilde{\Delta}_0[2]|)$ is:

$$P(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0)P(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} < 0) + P(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} < 0)P(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0).$$

We can thus rewrite, $P(|\tilde{\Delta}_0[1]| > |\tilde{\Delta}_0[2]|) - P(|\tilde{\Delta}_0[1]| < |\tilde{\Delta}_0[2]|) = P(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0)(P(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0) - P(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} < 0)) + P(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} < 0)(P(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} < 0) - P(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0))$. This expression is equivalent to $(2P(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0) - 1)(P(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0) - P(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} < 0))$. Since $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}$ has positive mean (recall that $a_0^{\{1\}} - a_0^{\{2\}} > 0$) and the difference of independent symmetric random variables is symmetric, then $P(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0) > 0.5$.[14] The same argument holds for $\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}}$ and thus $P(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0) > 0.5$. This implies that the expression above is positive. Since $P(|\tilde{\Delta}_0[1]| > |\tilde{\Delta}_0[2]|) = 1 - P(|\tilde{\Delta}_0[1]| < |\tilde{\Delta}_0[2]|)$, disagreement across attribute 1 is the most likely. In the extreme case in which $P(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0) = P(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0) = 1$ the, probability that disagreement takes place across attribute 1 is exactly one.

Let us consider now the case in which $\Delta_0[1] = \Delta_0[2] \geq 0$. We have that $\Delta_0[1] - \Delta_0[2] = a_0^{\{1\}} - a_0^{\{2\}} = 0$. Also, since differences are non-negative, $a_0^{\{1,2\}} - a_0^{\{\emptyset\}} \geq 0$ has to hold. This implies, again by symmetry, that $P(|\tilde{\Delta}_0[1]| > |\tilde{\Delta}_0[2]|) - P(|\tilde{\Delta}_0[1]| < |\tilde{\Delta}_0[2]|) = 0$. In this case, disagreement across either attribute is equally likely.

■

Proof of Remark 1. Suppose that disagreement persists across attribute 1. Notice that the expression of eventual attitudes is based on the expression of the deterministic ones in the main Theorem of chapter 1, that is, $a_\infty^A = 2^{-1}\tau(W^\infty)\Delta_0[1]$ if $1 \in A$ and $a_\infty^A = -2^{-1}\tau(W^\infty)\Delta_0[1]$ if $1 \notin A$, with $\tau(W^\infty) = 1 - \Delta_0[2]/\Delta_0[1]$. That is so because both, the homophily values and the differences in attitudes, preserve their

[13] Notice that as $\tilde{a}_0^{\{2\}}$ is continuous, so is $-\tilde{a}_0^{\{2\}}$. The sum $\tilde{a}_0^{\{1\}} + (-\tilde{a}_0^{\{2\}})$ is thus continuous. See Sheldon et al. (2002).

[14] See Stroock (2010).

properties when these differences enter in absolute value in the Luce form. Specifically, for every realization of initial attitudes and for every attribute i, $\Delta_t[i] = \lambda^i_{t-1}\Delta_{t-1}[i]$. Thus, $\Delta_t[i] \neq 0$ if and only if $\lambda^i_{t-1} \neq 0$ and $\Delta_{t-1}[i] \neq 0$. Also, given the Luce form, $\lambda^i_t \neq 0$ if and only if $\Delta_t[i] \neq 0$. Furthermore, $\Delta_0[i] > (<)\ 0$ implies that at every time t, $\Delta_t[i] > (<)\ 0$ and $\lambda^i_0 > 0$ implies that at every time t, $\lambda^i_t > 0$. Also, $\Delta_0[i] = 0$ and $\lambda^i_0 = 0$, imply that at every t, $\Delta_t[i] = 0$ and $\lambda^i_t = 0$, respectively.[15] For every realization of initial attitudes the ergodicity coefficient, $\tau(W^\infty)$, now becomes $1 - |\Delta_0[2]|/|\Delta_0[1]|$.[16] We thus have that $\tilde{a}^A_\infty = \overline{a}_0 + 2^{-1}(1 - |\tilde{\Delta}_0[2]|/|\tilde{\Delta}_0[1]|)\tilde{\Delta}_0[1]$ if $1 \in A$ and $\tilde{a}^A_\infty = \overline{a}_0 - 2^{-1}(1 - |\tilde{\Delta}_0[2]|/|\tilde{\Delta}_0[1]|)\tilde{\Delta}_0[1]$ if $1 \notin A$.[17]

■

Proof of Remark 2. It follows a similar reasoning as the proof of Remark 1. We therefore omit it here. ■

Proof of Proposition 2. Before proving the persistence of disagreement we state useful facts regarding the evolution of homophily values.

Let us focus first on the case in which $\Delta_0[1] > \Delta_0[2]$. It then follows that $\lambda^1_0 > 2^{-1} > \lambda^2_0$. Let us denote $\lambda^*_t = (\lambda^1_t + \beta^1_t)2^{-1} \in (0,1]$. Notice that as $\Delta_t[1] = 2^{-1}((a^{\{1,2\}}_t + a^{\{1\}}_t) - (a^{\{2\}}_t + a^{\{\emptyset\}}_t))$, using W_t in the main body, we can rewrite $\Delta_t[1] = \lambda^*_{t-1}\Delta_{t-1}[1]$. Similarly, we have that $\Delta_t[2] = (1 - \lambda^*_{t-1})\Delta_{t-1}[2]$. As these expressions hold for every t, we get that $\Delta_t[1] = \prod_{s=0}^{t-1} \lambda^*_s \Delta_0[1]$ and $\Delta_t[2] = \prod_{s=0}^{t-1}(1 - \lambda^*_s)\Delta_0[2]$. Now, for the ease of exposition let $\hat{\lambda}^1_t$, be the homophily value of attribute 1 under the process described in chapter 1, where W_t is symmetric. Now, since by assumption $\beta^1_t > \lambda^1_t$ at every t, then $\lambda^1_1 = \dfrac{\lambda^*_0\Delta_0[1]}{\lambda^*_0\Delta_0[1] + (1-\lambda^*_0)\Delta_0[2]} > \hat{\lambda}^1_1$. Thus $\Delta_2[1] = \lambda^*_1\Delta_1[1]$ is higher than in the symmetric case and, $\Delta_2[2]$ smaller. As a consequence, $\lambda^1_2 > \hat{\lambda}^1_2$. In general at every time t, $\beta^1_t > \lambda^1_t > \hat{\lambda}^1_t$ and $\beta^2_t < \lambda^2_t < \tilde{\lambda}^2_t$, with $\lambda^1_0 = \hat{\lambda}^1_0$. Consider now a sequence of ones. Since at every time t, $1 > \lambda^1_t \geq \hat{\lambda}_t$ and by step 7 in the proof of the main Theorem in chapter 1, $\lim_{t\to\infty} \hat{\lambda}^1_t = 1$, then $\lim_{t\to\infty} \lambda^1_t = 1$. Also since $1 \geq \beta^1_t > \lambda^1_t$, thus $\lim_{t\to\infty} \beta^1_t = 1$.[18] Recall that, by the same step, $\sum_{t=0}^\infty |log(\hat{\lambda}^1_t)|$ was convergent. Since at every t, $\lambda^*_t = (\lambda^1_t + \beta^1_t)2^{-1} > \hat{\lambda}^1_t$ then $|log(\lambda^*_t)| \leq |log(\hat{\lambda}^1_t)|$. Thus, by comparison $\sum_{t=0}^\infty |log(\lambda^*_t)|$ converges and hence $\prod_{t=0}^\infty \lambda^*_t = \delta \in (0,1]$. Since at every time t, $\lambda^*_t > \hat{\lambda}^1_t$, then $\delta > \mu = \prod_{t=0}^\infty \hat{\lambda}^1_t \in (0,1]$. Finally, since $\lim_{t\to\infty} 1 - \lambda^*_t = 0$, then $\prod_{t=0}^\infty (1 - \lambda^*_t) = 0$ holds.

We now proceed to state the persistence of disagreement. Let us consider the evolution on attitudes, given by $a_{t+1} = W_t a_t$. Notice that we can rewrite W_t as:

[15] For more details, see steps 1 and 7 in the proof of the main Theorem in chapter 1.

[16] See chapter 1, section 1.3.

[17] In contrast with chapter 1 we do not impose here that for every realization of initial attitudes, their average is equal to zero.

[18] Since $\beta^1_t > \lambda^1_t > \hat{\lambda}_t$, β^1_t and λ^1_t are closer to 1, their limiting value, at every time t. That means that the process converges faster to the eventual disagreement than in the symmetric case.

$$W_t = \frac{1}{4}\begin{array}{c} \begin{array}{cccc} \{1,2\} & \{1\} & \{2\} & \{\emptyset\} \end{array} \\ \begin{bmatrix} 1+\lambda_t^1+\lambda_t^2 & 1+\lambda_t^1-\lambda_t^2 & 1-\lambda_t^1+\lambda_t^2 & 1-\lambda_t^1-\lambda_t^2 \\ 1+\beta_t^1-\beta_t^2 & 1+\beta_t^1+\beta_t^2 & 1-\beta_t^1-\beta_t^2 & 1-\beta_t^1+\beta_t^2 \\ 1-\lambda_t^1+\lambda_t^2 & 1-\lambda_t^1-\lambda_t^2 & 1+\lambda_t^1+\lambda_t^2 & 1+\lambda_t^1-\lambda_t^2 \\ 1-\beta_t^1-\beta_t^2 & 1-\beta_t^1+\beta_t^2 & 1+\beta_t^1-\beta_t^2 & 1+\beta_t^1+\beta_t^2 \end{bmatrix} \end{array} \begin{array}{c} \\ \{1,2\} \\ \{1\} \\ \{2\} \\ \{\emptyset\} \end{array}$$

It then follows that $a_{t+1}^A = \overline{a}_t + 2^{-1}((-1)^{1+\mathbf{1}_i}\lambda_t^1\Delta_t[1] + \lambda_t^2\Delta_t[2])$ if $2 \in A$ and $a_{t+1}^A = \overline{a}_t + 2^{-1}((-1)^{1+\mathbf{1}_i}\beta_t^1\Delta_t[1] - \beta_t^2\Delta_t[2])$ if $2 \notin A$, where $\mathbf{1}_i$ is the indicator of type A possessing attribute 1. To make this point clearer, notice that using the ones in every entry in a row of matrix W_t multiplied by a_t, we compute average attitudes at $t+1$. Similarly, using the weights associated to attribute 1 (respectively attribute 2), i.e., λ_t^1 or β_t^1 (respectively λ_t^2 or β_t^2) we compute differences in average attitudes associated to attribute 1 (respectively attribute 2). They enter with positive sign if and only if the type possesses attribute 1 (respectively attribute 2). Using the expressions for $\Delta_t[1]$ and $\Delta_t[2]$ above we rewrite $a_{t+1}^A = \overline{a}_t + 2^{-1}((-1)^{1+\mathbf{1}_i}\lambda_t^1\prod_{s=0}^{t-1}\lambda_s^*\Delta_0[1] + \lambda_t^2\prod_{s=0}^{t-1}(1-\lambda_s^*)\Delta_0[2])$ if $2 \in A$ and $a_{t+1}^A = \overline{a}_t + 2^{-1}((-1)^{1+\mathbf{1}_i}\beta_t^1\prod_{s=0}^{t-1}\lambda_s^*\Delta_0[1] - \beta_t^2\prod_{s=0}^{t-1}(1-\lambda_s^*)\Delta_0[2])$ if $2 \notin A$. Since $\prod_{t=0}^{\infty}\lambda_t^* = \delta$, $\prod_{t=0}^{\infty}(1-\lambda_t^*) = 0$, $\lim_{t\to\infty}\beta_t^1 = 1$ and $\lim_{t\to\infty}\lambda_t^1 = 1$ it follows that $\lim_{t\to\infty} a_{t+1}^A = \lim_{t\to\infty}\overline{a}_t + 2^{-1}\delta\Delta_0[1]$ if $1 \in A$ and $\lim_{t\to\infty} a_{t+1}^A = \lim_{t\to\infty}\overline{a}_t - 2^{-1}\delta\Delta_0[1]$ if $1 \notin A$, provided that $\lim_{t\to\infty}\overline{a}_t$ exists. We prove that this is, in fact, the case. In doing so, notice that simple algebra yields $\overline{a}_t = \overline{a}_0 + 4^{-1}\sum_{s=0}^{t-1}(\lambda_s^2 - \beta_s^2)\Delta_s[2] = \overline{a}_0 + 4^{-1}\sum_{s=0}^{t-1}(\lambda_s^2-\beta_s^2)\prod_{m=0}^{s-1}(1-\lambda_m^*)\Delta_0[2]$. Thus, we need to prove that $\sum_{s=0}^{\infty}(\lambda_s^2-\beta_s^2)\prod_{m=0}^{s-1}(1-\lambda_s^*)$ exists. In doing so we prove that $\sum_{s=0}^{\infty}(\lambda_s^2-\beta_s^2)$ exists. Given that, at every time s, $\lambda_s^2-\beta_s^2 > (\lambda_s^2-\beta_s^2)\prod_{m=0}^{s-1}(1-\lambda_m^*)$ we will conclude, by comparison, that $\sum_{s=0}^{\infty}(\lambda_s^2-\beta_s^2)\prod_{m=0}^{s-1}(1-\lambda_m^*)$ exists. We proceed as follows: that $\sum_{s=0}^{\infty}(\lambda_s^2-\beta_s^2)$ diverges, is equivalent to state that $\prod_{s=0}^{\infty}(1-(\lambda_s^2-\beta_s^2)) = 0$.[19] We rewrite this expression as $\prod_{s=0}^{\infty}((1-\lambda_s^2)+\beta_s^2) = \prod_{s=0}^{\infty}(\lambda_s^1+\beta_s^2)$. From step 7 in the proof of the main theorem in chapter 1, we have that $\prod_{s=0}^{\infty}\hat{\lambda}_s^1 \in (0,1]$. Recall that since, at every s, $\lambda_s^1 \geq \hat{\lambda}_s^1$ and $\lambda_s^1+\beta_s^2 < 1$, it has to be that $\prod_{s=0}^{\infty}(\beta_s^2+\lambda_s^1) \in (0,1]$. Thus, $\sum_{s=0}^{\infty}(\lambda_s^2-\beta_s^2)$ exists, so does $\sum_{s=0}^{\infty}(\lambda_s^2-\beta_s^2)\prod_{m=0}^{s-1}(1-\lambda_m^*)$, by comparison. Now, Let $\alpha \in [0,\infty)$ be the value of this infinite sum. We thus have that $\lim_{t\to\infty} a_t^A = \overline{a}_0 + 4^{-1}\alpha\Delta_0[2] + 2^{-1}\delta\Delta_0[1]$ if $1 \in A$ and $\lim_{t\to\infty} a_t^A = \overline{a}_0 + 4^{-1}\alpha\Delta_0[2] - 2^{-1}\delta\Delta_0[1]$ if $1 \notin A$. As stated above, $\delta > \mu$, and thus, eventual disagreement is higher than in the symmetric case.

Let us now consider the case in which $\Delta_0[1] = \Delta_0[2]$. Then $\hat{\lambda}_t^1 = 2^{-1}$ at every time t, by the proof of the main Theorem in chapter 1. By the same arguments as above $\lambda_1^0 = \hat{\lambda}_1^0$ and $\lambda_t^1 > \hat{\lambda}_t^1$ at every time $t > 1$. Thus, $\Delta_t[1]$ is higher than the one in the symmetric case at every time $t > 1$. If we consider the process as starting at time $t = 1$ with $\lambda_t^1 > \hat{\lambda}_t^1$, same arguments as above follow. Thus, disagreement persists across attribute 1 and is higher than in the symmetric case. ■

[19] See Apostol (1977), chapter 8.

Proof of Proposition 3. This proof is based on the proof of the main Theorem in chapter 1. We proceed to explain, one in a row, which of its steps hold here. Step 1 describes a property that relies on both, the linearity of the updating process and on the Luce form. Specifically, Luce guarantees that at every t, $\Delta_t[i] \geq 0$ implies that $\lambda_t^i \geq 0$. By **(WDM)**, this step holds. Steps 2-5, dealing with the diagonalization of W_t, do not depend on the Luce form, we thus, they apply here. Step 6 only relies on non-negativity of homophily values and on the fact that $\sum_i \gamma_t^i = 1$, not on their specific form, thus it also holds. Step 7 absolutely relies on the Luce form. The new step is as follows:

Step 7. We first consider the case in which the limiting product of the homophily values is different from zero for one attribute i. Notice that this limiting product cannot be different from zero for more than one attribute. The reason is that in this case $\sum_i \lim_{t\longrightarrow\infty} \gamma_t^i > 1$ contradicting the properties of γ_t^i. We prove that $\prod_{t=0}^{\infty} \gamma_t^i = \mu^i$ with $\mu^i \in (0,1)$ for attribute i implies that $\prod_{t=0}^{\infty} \gamma_t^j = 0$ for attributes $j \neq i$. We also show that if there is such an attribute, it has to be attribute 1. Second, we consider the case in which the limiting product of homophily values is zero for all attributes. Before proceeding, recall that $\lim_{t\longrightarrow\infty} \gamma_t^i$ exists for every attribute i and $\lim_{t\longrightarrow\infty} \sum_i \gamma_t^i = \sum_i \lim_{t\longrightarrow\infty} \gamma_t^i = 1$.

First, suppose that $\prod_{t=0}^{\infty} \gamma_t^i = \mu^i$ with $\mu^i \in (0,1)$, or equivalently, that $\sum_{t=0}^{\infty} log(\gamma_t^i)$ exists. This implies that $\lim_{t\longrightarrow\infty} \gamma_t^i = 1$. Thus, $\lim_{t\longrightarrow\infty} \gamma_t^j = 0$ for every attribute j, implying that for non of them, $\prod_{t=0}^{\infty} \gamma_t^j = \mu^j$ with $\mu^j \in (0,1)$, but $\prod_{t=0}^{\infty} \gamma_t^j = 0$. Now, recall that by **(ADM)** and step 1 in the proof of the main Theorem in chapter 1, at every t, $\Delta_t[1] \geq \Delta_t[2] \geq \cdots \geq \Delta_t[n] \geq 0$ and $\gamma_t^1 \geq \gamma_t^2 \geq \cdots \geq \gamma_t^n \geq 0$ hold. Suppose, that there is an attribute $i > 1$ for which $\prod_{t=0}^{\infty} \gamma_t^i = \mu^i$ with $\mu^i \in (0,1)$ holds. This implies that $\lim_{t\longrightarrow\infty} \gamma_t^i = 1$ and $\lim_{t\longrightarrow\infty} \gamma_t^1 = 0$. Thus, for high enough t, γ_t^i would be arbitrarily close to 1 while γ_t^1 would be arbitrarily close to 0. Given that, at every t, $\gamma_t^1 \geq \gamma_t^2 \geq \cdots \geq \gamma_t^n \geq 0$ holds, the former statement cannot be true. We therefore conclude that $\prod_{t=0}^{\infty} \gamma_t^1 = \mu^1$ with $\mu^1 \in (0,1)$ and $\prod_{t=0}^{\infty} \gamma_t^i = 0$ for attributes $i > 1$. Under this scenario we have that:

$$W^{\infty} = \frac{1}{2^n} \begin{bmatrix} 1+\mu^1 & \cdots & 1+\mu^1 & 1-\mu^1 & \cdots & 1-\mu^1 \\ 1+\mu^1 & \cdots & 1+\mu^1 & 1-\mu^1 & \cdots & 1-\mu^1 \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 1+\mu^1 & \cdots & 1+\mu^1 & 1-\mu^1 & \cdots & 1-\mu^1 \\ 1+\mu^1 & \cdots & 1+\mu^1 & 1-\mu^1 & \cdots & 1-\mu^1 \\ 1-\mu^1 & \cdots & 1-\mu^1 & 1+\mu^1 & \cdots & 1+\mu^1 \\ 1-\mu^1 & \cdots & 1-\mu^1 & 1+\mu^1 & \cdots & 1+\mu^1 \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 1-\mu^1 & \cdots & 1-\mu^1 & 1+\mu^1 & \cdots & 1+\mu^1 \\ 1-\mu^1 & \cdots & 1-\mu^1 & 1+\mu^1 & \cdots & 1+\mu^1 \end{bmatrix},$$

where $\mu^1 = \prod_{t=0}^{\infty} \gamma_t^1$.

Second, suppose that $\prod_{t=0}^{\infty} \gamma_t^1 = 0$ and $\lim_{t\longrightarrow\infty} \gamma_t^1 = 1$. It implies that $\lim_{t\longrightarrow\infty} \gamma_t^j = 0$ for every attribute $i > 1$. Thus, no attribute $i > 1$ is such that $\prod_{t=0}^{\infty} \gamma_t^i = \mu^i$ with $\mu^i \in (0,1)$. Suppose now that $\prod_{t=0}^{\infty} \gamma_t^1 = 0$ and $\lim_{t\longrightarrow\infty} \gamma_t^1 = \alpha$ with $\alpha \in (0,1)$. Then, $\sum_{i>1} \lim_{t\longrightarrow\infty} \gamma_t^i = 1 - \alpha$ with $1 - \alpha \in (0,1)$. Thus, no attribute i is such that $\prod_{t=0}^{\infty} \gamma_t^i = \mu^i$ with $\mu^i \in (0,1)$. Finally, suppose that $\prod_{t=0}^{\infty} \gamma_t^1 = 0$ and $\lim_{t\longrightarrow\infty} \gamma_t^1 = 0$. Then, either for exactly one attribute $i > 1$, $\lim_{t\longrightarrow\infty} \gamma_t^i = 1$, or for some (possibly all) attributes $i > 1$, $\sum_{i>1} \lim_{t\longrightarrow\infty} \gamma_t^i = 1$. None of these cases can hold. The reason is that for high enough t, some γ_t^i would be arbitrarily close to a positive number (which is 1 when for exactly one attribute $i > 1$, $\lim_{t\longrightarrow\infty} \gamma_t^i = 1$) while γ_t^1 would be arbitrarily close to 0. Since, at every t, $\gamma_t^1 \geq \gamma_t^2 \geq \cdots \geq \gamma_t^n \geq 0$ holds, this cannot be true. We therefore conclude that in all these cases $\prod_{t=0}^{\infty} \gamma_t^i = 0$ for all attributes. Under this scenario all entries of W^{∞} are $(2^n)^{-1}$.

Step 8. By step 7, the necessary and sufficient condition for disagreement to persist is that $\sum_{t=0}^{\infty} log \gamma_t^1$ exists. When this is the case, eventual attitudes take the same form as in step 8 in the proof of the main Theorem in chapter 1, where $\tau(W^{\infty}) = \lim_{T\rightarrow\infty} \prod_{t=0}^{T} \gamma_t^1$.

Step 9. It depends partially on Luce, since it guarantees that, at every t, $\Delta_t[i] \geq 0$ implies that $\lambda_t[i] \geq 0$. By **(WDM)** this step holds. ■

Chapter 3

When to Do the Hard Stuff? Dispositions, Motivation and the Choice of Difficulties

3.1 Introduction

One of the reasons as to why individuals tend to avoid difficult tasks is because they do not feel able enough to confront them. Not coping with them might, however, imply foregoing the opportunity of getting better economic outcomes, not available otherwise. As Liebow (1967) documents in his study of the Negro male community of Washington inner city:

Convinced of their inadequacies, not only do they not seek out those few better-paying jobs which test their resources, but actively avoid them, gravitating in a mass to the menial, routine jobs which offer no challenge -and therefore posse not threat-to the already diminished images they have of themselves(...). Thus, the man's low self-esteem generates a fear of being tested and prevents him from accepting a job with responsibilities or, once on a job, form staying with it if responsibilities are thrust on him, even if wages are commensurably higher.

The story above offers two interesting insights. The first one is that individual dispositions might dramatically influence decisions of huge economic relevance. When documenting the relationship between the *achievement motive*, that is, individual dispositions to strive for success, and upward mobility patterns in the United States, Atkinson and Feather (1966) highlight how, despite of the fact that education is the main determinant of upward mobility, individual dispositions should not be neglected. In fact, 65% of the people who exhibited upward mobility patterns at a higher extent, only had high school education or less.[1] The second one is the trade-off between tasks' difficulty and their economic outcomes. While a routine job is probably more easily developed than a very demanding one, good economic outcomes, as higher wages or promotion opportunities, might only be available in the latter.[2]

Our purpose in this paper is to understand and highlight the role played by individual dispositions in shaping avoidance behavior. We interpret individual dispositions as an expression of non-cognitive abilities.[3] Examples of non-cognitive abilities are emotional stability, that manifests, among others aspects, in self-confidence and self-esteem, or conscientiousness, that manifests, among others aspects, in perseverance.[4] In order to do it, we develop a tractable model in which the decision maker, henceforth DM, who is characterized by a disposition (that is, a non-cognitive abilities level), decides the optimal time to deal with difficulties.

Our approach is as follows. We consider a dynamic framework in which at every point in time the decision maker, henceforth DM, might experience two states, namely, the full capacity state and the deteriorated capacity state, with a constant probability. In the full capacity state she enjoys high dispositions while in the deteriorated capacity state her dispositions are low. Tasks are of two types, namely, easy and difficult. On the one hand, getting good economic outcomes is less likely under difficult than under easy tasks but on the other hand, outcomes associated to difficult tasks are more valuable than outcomes associated to easy tasks. We consider that states and economic performance are positively related, specifically, the higher the DM's disposition the higher the probability of being successful when developing a task, either easy or difficult. It is worth mentioning that no effort decision is analyzed here. The only decision of the DM has to make is when to confront difficulties. We assume that once she decides to confront them, she sticks at this decision forever.

We also study the case in which the DM's disposition is sensitive to outcome achievements. As Mruk (2006) points out, the demands of life are not constant, so self-esteem levels will fluctuate depending on what is happening in a person's life. Redundancy, bereavement, illness, studying, gaining a qualification, parenthood, poverty, being a victim of crime, divorce, promotion at work will all have an impact on our self-esteem levels. Self-esteem levels go up and down and can change over time. Also,

[1] To establish comparisons between the prestige of occupations of parents and sons, private households populated by people older than 21 were interviewed. Specific measures related to individual dispositions to strive for success were collected.

[2] Also, as Atkinson and Feather (1966) suggest, high prestige occupations are perceived as being more difficult to attain than low prestige occupations. This hierarchy can be seen as a series of tasks in which the outcome value comes together with difficulty.

[3] We will interchangeably use the term state, disposition or simply ability when referring to the non-cognitive ability level that the individual enjoys.

[4] See John and Srivastava (1999) for the Big Five domains of non-cognitive abilities, their traits and facets.

as Bénabou and Tirole (2002) point out, motivation helps individuals to persevere in the presence of setbacks. We formalize this idea by allowing the probabilities of experiencing the full capacity state and the deteriorated capacity state to evolve over time. Specifically, we assume that their value at a given period depends on their value and on the likelihood of good economic outcomes in the previous period.[5]

Our results are as follows. We find how a low disposition DM will avoid difficulties forever while a high disposition DM will cope with them since the beginning. Thus, individuals with poor abilities get trapped into low value easy tasks. However, when motivation plays a role, the achievement of good economic outcomes out of easy tasks leaves the DM with the disposition of coping with difficulties from some point in time on.[6] In line with this finding it is worth mentioning the results of a program carried out in West Bengal, by the indian microfinance institution *Bandhan*, consisting on providing extremely poor individuals with productive assets. The authors observed how people ended up working 28% more hours, mostly on activities not related to the assets they were given and that their mental health had improved. The program was considered to have injected a dose of motivation, that pushed people to start new economic activities.[7]

Our proposal is closely related to the branch of literature that links poverty and psychology. For instance, Dalton et al. (2014) discuss the importance of aspirations failure in the perpetuation of poverty. This paper, as ours, highlights the role of internal constraints as a source of behaviors that might preclude individuals from getting high welfare achievements. Their research question is, however, different from ours, whereas they focus in one particular bias, namely, aspiration failure, we analyze the role of non-cognitive abilities. We also find relations with the literature of addiction and self-control. Specifically, Bernheim and Rangel (2004) study patterns of addictive behavior of a DM that operates in two modes, namely, cold and hot. When in the cold mode, the DM selects her most preferred alternative whereas when in the hot mode, choices and preferences may diverge because the DM losses cognitive control. This paper presents a theory of addiction whereas ours focuses on the effects of non-cognitive abilities in the decision of facing onerous but valuable tasks. Also, Ozdenoren et al. (2012) exhaustively account for the dynamics of self-control performance of a DM that has to choose her optimal consumption path. We depart from this paper since we focus on outcome achievement and motivation, and not on capacity exhaustion, as the main driver of decisions.

The paper is organized as follows. Section 3.2 presents the baseline model. In this

[5]In particular we assume that probabilities evolve according to a Markov process.

[6]This is consistent with Ali (2011), a model in which a long-run self, the planner, has to decide, at every point in time, whether to allow the short-run self, the doer, to face a menu in which a tempting alternative is available. The planner does so whenever the doer experiences high self-control.

[7]The full article is available at http://www.economist.com/node/21554506.

version, the probability of experiencing the full capacity state is time independent. Section 3.3 presents the extended model. In it, the probability of experiencing the full capacity state is sensitive to outcome achievements. The dynamics of its evolution is therefore outlined. In both sections, optimal strategies and their associated utility gains are presented. Section 3.4 concludes. Section 3.5 contains the technical proofs.

3.2 A model on avoidance behavior

Let s_1 and s_2 denote the two states that the DM might experience. When experiencing s_2 the DM is in the full capacity state and enjoys high abilities. When experiencing s_1 the DM is in the deteriorated capacity state, meaning that she executes her abilities poorly. At every point in time, $t \in \mathbb{Z}_+$, she has a probability $q \in [0, 1]$ of experiencing s_2. Thus, she experiences s_1 with probability $1 - q$. Tasks are of two types, easy or the difficulty level d_1, and difficult or the difficulty level d_2.

The likelihood of getting good economic outcomes is denoted p_{ij}, with $i, j = 1, 2$. Subscript i refers to the DM's state, that is, either s_1 or s_2, whereas subscript j refers to the difficulty of the task, that is, either d_1 or d_2. Probabilities are as follows: first, fixing difficulty, the likelihood of good economic outcomes increases with the DM's state. There is, in fact, a large amount of literature posing non-cognitive abilities as one of the factors determining performance and outcomes, for instance, in the domains of education and in the labor market.[8] Second, fixing the DM's state, the likelihood of good economic outcomes decreases with task's difficulty. The following table presents these probabilities:

Table 1. Success probabilities

	s_1		s_2
d_1	p_{11}	$<$	p_{21}
	$\vee$		$\vee$
d_2	p_{12}	$<$	p_{22}

Notice that no direct relation is established between p_{11} and p_{22}. Finally, good economic outcomes are worth just 1 unit when they are the result of developing easy tasks and $K > 1$ units when they come out of developing difficult tasks.

We make an assumption regarding the success probabilities. It captures the idea that individuals with low dispositions are more vulnerable than individuals with high dispositions to the characteristics of the tasks they deal with. For high disposition

[8]See Heckman et al. (2006) and Balart and Cabrales (2014) for an illustration of this relationship.

individuals, task's difficulty is less relevant than for low disposition individuals in determining their chances of success.[9] We formally express it as:

Assumption 1: $p_{11} - p_{12} > p_{21} - p_{22}$.

The second assumption is related to the strategies among which the DM chooses:

Assumption 2: once the DM decides to face difficulties, she commits to this decision forever.

In fact, there exists evidence showing that in many situations individuals become locked into (possibly) costly courses of action and a cycle of escalation of commitment arises. The justification of previous decisions, the necessity to comply with norms or a desire for decision consistency in the decision making process, might encourage commitment.[10] The strategies available to the DM therefore comprise choosing the point in time in which to face difficulties. We denote by (∞) the *always avoiding difficulties* strategy and by (0) the *facing difficulties since the beginning* strategy. A strategy consisting on facing difficulties from a point in time $0 < t < \infty$ on, is denoted (t).[11]

The DM behaves as an expected utility maximizer. Thus, she determines her optimal path of action at the initial point in time, taking into account her disposition, that is, the point-wise probability q of being in the full capacity state. We consider that the DM is risk neutral. We then focus on the role of dispositions without dealing with risk aversion issues.[12] Thus, the current expected utility of developing an easy task at time t is $qp_{21} + (1-q)p_{11}$ and that of developing a difficult task is $K(qp_{22} + (1-q)p_{12})$. Furthermore, let $\delta \in (0, 1)$ denote the discount factor of the stream of pay-offs. We formally state the DM's problem as follows:

When experiencing the full capacity state, s_2, with probability q, the DM decides, at $t = 0$, the point in time t to face difficult tasks, in order to maximize her long-run expected utility. Specifically, she solves:

[9] On the domain of cognitive abilities Gonzalez (2005) provides experimental evidence on how increasing task difficulty, understood as workload, was more detrimental for low ability individuals.

[10] See Staw (1981) for the concept of escalation of commitment. See also Arkes and Blumer (1985) and Thaler (1980) for a justification of this phenomenon based on the sunk cost effect.

[11] It is worth to stress how the explicit introduction of time does not aim to describe the evolution of abilities along the life cycle. Time only aim to capture the point-wise choice of task difficulty.

[12] See Tanaka et al. (2010) for a paper studying the relationship between poverty and risk and time preferences.

$$\underset{t}{Max}\ u((t)) = \underset{t}{Max} \sum_{i=0}^{t-1} \delta^i (qp_{21} + (1-q)p_{11}) + K \sum_{i=t}^{\infty} \delta^i (qp_{22} + (1-q)p_{12}).$$

As stated, the only decision the DM has to make is when to jump into difficulties. The second part of the sum above reflects the fact that once she decides to do so, she sticks at this decision forever.

3.2.1 Results

It seems intuitive that individuals who enjoy better dispositions tend to perform tasks better. In fact, it is common that people tend to avoid difficulties when they do not feel prepared to face them. The results in this section capture this idea. When the DM experiences the full capacity state with high enough probability, she will opt for difficulties since the beginning. In contrast, when the probability of experiencing the full capacity state is low enough, she will prefer to avoid them forever.

We build our results building a function $\lambda : [0,1]^4 \times \mathbb{R} \longrightarrow \mathbb{R}$, that depends on the primitives of the model, defining the environment in which the DM makes her decision, namely, the success probabilities and the value of these outcomes. It defines a domination threshold between the strategy of facing difficulties since the beginning, (0), and the strategy of postponing them for one period (1).[13] For values of q higher or equal than this threshold, (0) is preferred to (1) and for values of q smaller than it, (1) is preferred to (0). This information will be enough to identify the optimal strategy.

Before stating the result it is worth highlighting that whenever outcomes out of difficult tasks do (respectively do not) compensate the decrease in probability of successfully dealing with them, that is, whenever $p_{11}/p_{12} \leq K$ (respectively $K \leq p_{21}/p_{22}$), the DM finds optimal to always face (respectively to always avoid) difficulties, even if she is of extreme low disposition, that is, if $q = 0$ (respectively, even if she is of extreme high disposition, that is, if $q = 1$). We then focus on the interesting case in which $p_{21}/p_{22} < K < p_{11}/p_{12}$. Results are as follows:

Theorem 1. *The DM's optimal strategy is to face difficulties since the beginning whenever she enjoys the full capacity state with high enough probability (that is, whenever $\lambda \leq q$) and to always avoid them whenever she experiences the full capacity state with low enough probability (that is, whenever $q < \lambda$).*

[13]Specifically, it is the result of equating the long-run expected utility of (0) and one of (1), under all possible q. It gives the us the q such that the DM will be indifferent between not postponing difficulties and doing so for one period. It value is $\lambda = \dfrac{Kp_{12} - p_{11}}{(p_{21} - p_{11}) - K(p_{22} - p_{12})}$. See the proof of Theorem 1.

Optimal paths of action are extreme behaviors. Facing difficulties from an intermediate point in time is never optimal. Notice that, as going back from difficult to easy tasks is never considered by the DM, assumption 2 is immaterial here. Notice also that if for the DM never (respectively always) postponing difficulties is optimal, this is also be the case for any DM with a higher (respectively lower) disposition. We interpret the always avoiding difficulties strategy as procrastination on onerous tasks.[14]

The following example aims to clarify the elements of the model and these results:

Example. Consider a DM who is deciding which type of job to look for or to accept. An easy (routine) job gives the DM a payoff (wage) of 1 whereas a difficult (high responsibility) job has payoff $K = 1.3$. The success probabilities in either job are:

	s_1	s_2
d_1	0.7	0.8
d_2	0.5	0.7

In this case $\lambda = 0.31$. Thus if the DM is of low enough disposition (that is, if $q < 0.31$), she finds optimal to always postpone the acceptance of the high responsibility job, whereas if she is of high enough disposition (that is, if $q \geq 0.31$), she will find optimal to deal with the high responsibility job since the beginning. We depict below the ranking of long-run expected utilities under the strategies the DM chooses among. The left figure illustrates the case in which always avoiding difficulties is optimal whereas the right figure illustrates the case in which facing them since the beginning is optimal:

Figure 1. The optimal strategy is (∞)

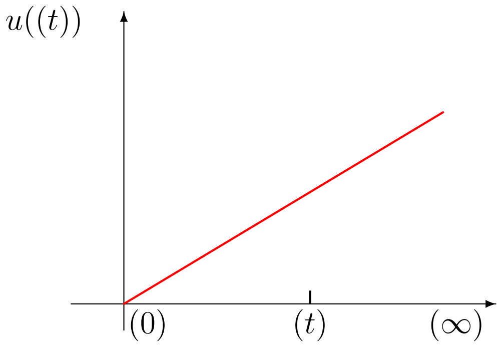
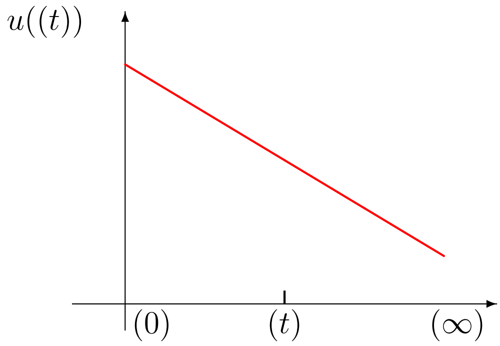


Figure 2. The optimal strategy is (0)



[14]See O'Donoghue and Rabin (2001) and O'Donoghue and Rabin (2008) for two references on procrastination.

The following remark discusses how the threshold λ reacts to the primitives of the model, defining the environment in which the DM make her decision, namely the success probabilities and the value of good economic outcomes under difficult tasks:

Remark. *The threshold λ is decreasing in p_{22} and p_{12} and K and increasing in p_{21} and p_{11}.*

Higher chances of successfully developing either easy or difficult tasks make the DM more prone to choose each of them. In the same vein, an increase in the value of good economic outcomes out of difficult tasks incentivizes the DM to face them. An increase in K directly raises the utility of facing difficult tasks. This happens despite of the fact that the probability of achieving good economic outcomes in these circumstances is systematically lower. The results above are summarized as follows:

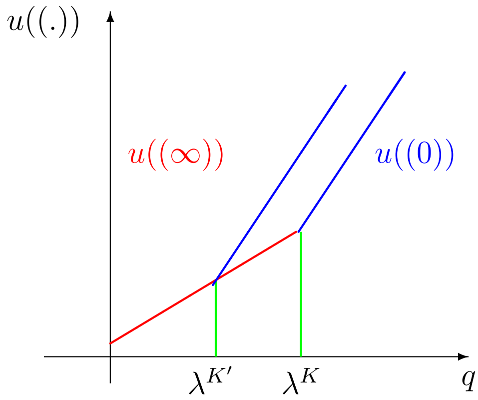
Corollary. *The range of probabilities $[q, 1]$ such that the DM's optimal strategy is to face difficulties since the beginning increases (respectively decreases) with the probability of success under difficult (respectively easy) tasks. It also increases with the value of economic outcomes out of difficult tasks.*

We also describe the utilities derived out of the aforementioned optimal strategies and analyze the effect of a marginal boost in dispositions. For this purpose we assume that a marginal increase in dispositions, does not affect the originally optimal strategy. Also, let us denote $\alpha \equiv \dfrac{1}{1-\delta}$. Results are as follows:

Proposition 1. *The long-run expected utility of any optimal strategy is monotonically increasing and linear in q. Its value is $\alpha K(qp_{22} + (1-q)p_{12})$ whenever the DM finds optimal to face difficulties since the beginning and $\alpha(qp_{21} + (1-q)p_{11})$ whenever the DM finds optimal to always avoid difficulties. Moreover, the marginal return of an increase in the DM's disposition is higher when the DM is already of high dispositions (that is, when $q \geq \lambda$) than when she is of low dispositions (that is, whenever $q < \lambda$).*

In the following figure we illustrate the aforementioned statements . In the x-axis we represent the DM dispositions, that is, the probability of experiencing the full capacity state. In the y-axis we represent the ranking of long-run expected utilities derived from the two possible optimal strategies, namely, either always avoiding difficulties or always facing them, according to the aforementioned probability. Observe how a marginal increase in the DM's disposition, increases the utility of any path of action and in particular, of the optimal one. Observe also how the threshold λ, which decreases with K, generates a kink, making utility convex (specifically, piece-wise linear) in q:

Figure 3. Utility as a function of dispositions, with $K' > K$



To conclude this section, we discuss on the possibility of carrying out a welfare assessment analysis. The intuition is as follows: consider two individuals. One of them, the disadvantaged individual, has low abilities and always avoids difficulties, the other, the advantaged individual, has high abilities and always faces difficulties. It turns out that the marginal return of boosting abilities is higher for the advantaged individual than for the disadvantaged individual. Suppose that a social planner has one unit of resources, devoted to improve abilities. If it is the case that the planner only cares about maximizing total returns, he might allocate this unit on the advantaged individual. If he also has equity concerns, he will have to take into account that the utility gap between the advantaged and the disadvantaged individual will exacerbate. In this case, the planner might be willing to allocate resources on the disadvantaged individual.

3.3 The role of motivation

In this section we model how successes and failures might affect the manifestation of non-cognitive abilities, for instance, self-esteem, self-confidence or perseverance.

We assume that the probability of experiencing the full capacity state varies over time according to a Markovian process. This modeling aims to capture the idea that success may boost the manifestation of the non-cognitive abilities while failure may deteriorate it. Formally, the probability of experiencing the full capacity state at time t, depends on its value at time $t-1$, and also on the success probabilities. Let $q^{(t)} \in [0, 1]$ denote the probability of experiencing the full capacity state.[15] Let the success probablities be the ones described in table 1. The following expression accounts for the evolution of the probabilities of experiencing either state:

[15]Thus, $1 - q^{(t)}$ denotes the probability of experiencing the deteriorated capacity state at time t.

$$[q^{(t-1)} \quad 1-q^{(t-1)}]\begin{bmatrix} p_{2j} & 1-p_{2j} \\ p_{1j} & 1-p_{1j} \end{bmatrix} = [q^{(t)} \quad 1-q^{(t)}], \tag{3.1}$$

where $j = \{1, 2\}$ accounts for task's difficulty.[16] Consider that at time $t-1$ the DM experiences s_2 with probability $q^{(t-1)}$. Then, at time t she will experience s_2 with the probability with which she was successful in the previous period. This is captured by the first column in the matrix above. Similarly, at t she will experience s_1 with the probability with which she failed in the previous period. This is captured by the second column the matrix above. Let $q^{(0)}$ be the DM's initial probability of experiencing the full capacity state. The current expected utility of developing an easy task at time t is $q^{(t)}p_{21} + (1-q^{(t)})p_{11}$ and the one developing a difficult task is $K(q^{(t)}p_{22} + (1-q^{(t)})p_{12})$.

Notice that, as a consequence of $q^{(t)}$ evolving according to a Markovian process, two stationary probabilities arise. These are, the one related to always facing easy tasks, denoted q^e, and the one related to always facing difficult tasks, denoted q^d. We interpret them as the average long-run frequencies with which the DM experiences the full capacity state when she always faces easy or difficult tasks, respectively. Since the likelihood of success is higher in easy tasks, it is the case that the DM is eventually better off in terms of capacities when she decides to only face easy tasks than when she decides to only face difficult tasks, more formally, $q^e > q^d$.[17] The DM's problem is as follows:

When experiencing the full capacity state, s_2, with probability $q^{(0)}$, she decides, at $t = 0$, the point in time t, to face difficult tasks, in order to maximize her long-run expected utility. Specifically, she solves:

$$\underset{t}{Max}\ u((t)) = \underset{t}{Max} \sum_{i=0}^{t-1} \delta^i (q^{(i)}p_{21} + (1-q^{(i)})p_{11}) + K \sum_{i=t}^{\infty} \delta^i (q^{(i)}p_{22} + (1-q^{(i)})p_{12}).$$

As stated, the only decision the DM has to make is when to jump into difficulties, in an environment in which her current performance is sensitive to previous outcomes achievement. The second part of the sum above reflects the fact that once she decides to do so, she sticks at this decision forever.

[16] Notice that $q^{(t)}$ depends on the chosen strategy. If the DM decides to face difficult tasks from $t = 5$ on, $q^{(4)}$ is the resulting probability of having faced easy tasks for four periods. If she decides to face difficult tasks from $t = 3$ on, $q^{(4)}$ is the resulting probability of having faced easy tasks for two periods and difficult ones from the third period on.

[17] Let $\mathbb{T}^k$, with $k = e, d$, denote the transition matrices out of facing either easy or difficult tasks, involved in expression (3.1), respectively. Their determinants are $T^e = p_{21} - p_{11}$ and $T^d = p_{22} - p_{12}$, respectively. In getting q^e and q^d we solve $[q^k, 1-q^k]\mathbb{T}^k = [q^k, 1-q^k]$. We have that $q^e = (p_{11})(1 - T^e)^{-1}$ and $q^d = (p_{12})(1-T^d)^{-1}$. Suppose that $q^e < q^d$. This implies that $p_{11}(1-T^d) < p_{12}(1-T^e)$ or $p_{11}(1-p_{22}) < p_{12}(1-p_{21})$ which cannot hold since $p_{11} > p_{12}$ and $(1-p_{22}) > (1-p_{21})$. Thus $q^e > q^d$ has to hold.

3.3.1 Results

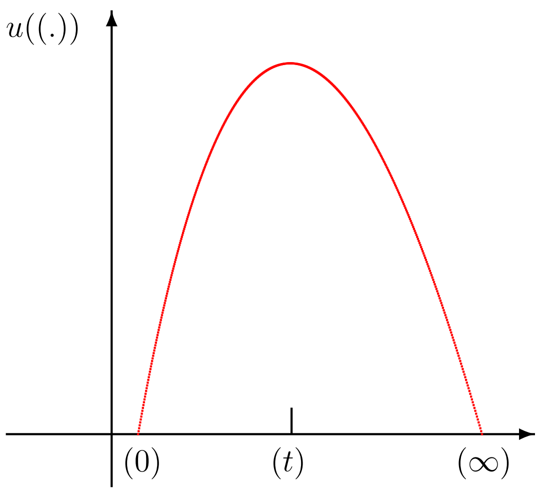
As in the previous section, we instrument our analysis using a function μ, depending on the primitives of the model. It defines a domination threshold between the strategy of facing difficulties since the beginning and the strategy of postponing them for one period, that is, between (0) and (1). For values of $q^{(0)}$ higher or equal than this threshold, (0) is preferred to (1) and for values of $q^{(0)}$ smaller than it, (1) is preferred to (0). This threshold and the stationary probabilities, determine optimal strategies. Let us first focus on the case in which assumption 2 does not play a role, that is, when DM's optimal strategy, in fact, belongs the class of strategies prescribed by assumption 2. We comment on the remaining cases afterwards. Results are as follows:

Theorem 2. *The DM's optimal strategy is to face difficulties since the beginning whenever she always enjoys the full capacity state with high enough probability (that is, whenever $\mu \leq q^d < q^e, q^{(0)}$), to always avoid them whenever she always experiences the full capacity state with low enough probability (that is, whenever $q^{(0)}, q^d < q^e \leq \mu$) and to face them from an intermediate point in time whenever she gets motivated through outcome achievements associated to easy tasks (that is, whenever $q^{(0)} < \mu \leq q^d < q^e$).*

In contrast with results in the previous section, jumping into difficult tasks at some point in time can be optimal here. We interpret this strategy as one in which the DM prefers to first deal with easy tasks, because performing properly motivates her to deal with difficult but more rewarding tasks.[18]

In the following figure we depict the ranking of utilities in this case. The DM exhibits single-peaked preferences on the optimal time to face difficulties, with the peak corresponding to an intermediate strategy:

Figure 4. The optimal strategy is (t)



[18] It is worth mentioning that if $q^d = q^{(0)}$ (respectively $q^e = q^{(0)}$) when facing (respectively avoiding) difficulties since the beginning is optimal, we are back to the optimal behavior in the previous section.

The question of when to do the hard stuff arised in Quora, an internet knowledge market, in which people discuss about a specific given topic. The topic was: Is it better to do easy tasks first and then move on to harder ones, or vice versa?[19] One of the answers, that accurately illustrates our statement, was:

Important is to evaluate, which are the harder tasks and which the easy tasks. Out of this it becomes clear, how long it will take to do them. (...) The rest has more psychological character and is strongly depending on the personality. I personally like to mix it. This gives the success feeling, if you do the easy tasks and motivates, to continue with the harder tasks, to make the overall project the success.

If individuals indeed behave this way, there will be chances of improving individual achievements through playing with motivation. A model of human capital accumulation in which individuals build their skills by developing easy tasks up to the point that it is optimal for them to face difficulties, might offer the same type of results. However we truly think that the human capital accumulation story is essentially different from the motivational story. This difference relies on the following reasoning: while individuals build their human capital in the actual process of developing a task, motivation results when outcomes are achieved. We think that this is a crucial distinction, that would posse different policy implications.

We now briefly comment on some cases in which assumption 2 plays a role. That is, cases in which the DM has to choose the optimal strategy among the class of strategies prescribed by assumption 2, regardless of whether other path of action would have delivered higher utility. Under $q^d < q^e < \mu \leq q^{(0)}$ the DM would have preferred to switch to easy tasks after have been dealing with difficulties for a while. Within the class of strategies she can choose among due to assumption 2, the DM exhibits single deep preferences on the optimal time to face difficulties. The deep corresponds to an intermediate strategy and the peaks correspond to the extreme strategies or either never dealing with difficulties or facing them since the beginning. The same happens under $q^d < \mu \leq q^e, q^{(0)}$. Among the available strategies prescribed by assumption 2, the DM ends up dealing with difficulties since the beginning. Finally under $q^{(0)}, q^d \leq \mu < q^e$ the DM would have also preferred to switch to easy tasks after have been dealing with difficulties for a while. As a result of assumption 2, the best thing the DM can do is to perform an intermediate strategy.[20]

The following result deals with the properties of the utility under the three optimal strategies. It also describes the returns of a boost in the DM's initial disposition, that

[19]See http://www.quora.com/Is-it-better-to-do-easy-tasks-first-and-then-move-on-to-harder-ones-or-vice-versa.

[20]See the proof of Theorem 2.

is, $q^{(0)}$. We assume that a marginal increase in the initial disposition, does not affect the originally optimal strategy. Specifically, for the case in which an intermediate strategy is optimal we consider that marginal increase in the initial disposition, does not affect the particular point in time to face difficulties.[21] Before stating the results let us denote by $mr((.))$, the marginal return of a increase in the DM's initial disposition, under any optimal strategy. Proposition 2 is as follows:

Proposition 2. *The long-run expected utility is monotonically increasing and linear in $q^{(0)}$ under any optimal strategy. Moreover, $mr((0)) > mr((t-1)) > mr((t)) > mr((\infty))$.*

These results, as the ones in Proposition 1, capture the idea that individuals with better abilities perform better and achieve higher utility. It is also the case that advantaged individuals, those with high $q^{(0)}$, benefit more from a marginal increase in their abilities. As the table below illustrates, as $q^{(0)}$ increases within a row, everything else equal, that is, as the DM is of higher initial dispositions, she moves from finding (∞) or (t) optimal to finding (0) optimal.

In the previous section we illustrated how the utility of high disposition individuals that always confront difficulties was higher than the utility of low disposition individuals that always avoid difficulties. We also carry such an analysis in this framework. As $q^e > q^d$, we list in the table below the three combinations in which μ, q^e and q^d may appear, the way in which $q^{(0)}$ relates to them and the optimal strategies. Within each combination we consider that μ, q^e and q^d remain unaltered. However, they might be different across combinations. For the ease of exposition we only consider strict inequalities:[22]

Table 2. Optimal strategies

C.1	$q^{(0)} < q^d < q^e < \mu$ (∞)	$q^d < q^{(0)} < q^e < \mu$ (∞)	$q^d < q^e < q^{(0)} < \mu$ (∞)	$q^d < q^e < \mu < q^{(0)}$ (0) or (∞)
C.2	$q^{(0)} < \mu < q^d < q^e$ (t)	$\mu < q^{(0)} < q^d < q^e$ (0)	$\mu < q^d < q^{(0)} < q^e$ (0)	$\mu < q^d < q^e < q^{(0)}$ (0)
C.3	$q^{(0)} < q^d < \mu < q^e$ (t)	$q^d < q^{(0)} < \mu < q^e$ (t)	$q^d < \mu < q^{(0)} < q^e$ (0)	$q^d < \mu < q^e < q^{(0)}$ (0)

Let us focus on optimal strategies in the row corresponding to C.2, that is, either (t) or (0). In this case assumption 2 does not play a role. This allows us to make a

[21]Formally, when (t) is optimal then $t|_{q^{(0)}} = t|_{q^{(0)}+\epsilon}$ holds.
[22]For the complete analysis, see the proof of Theorem 2

neat comparison of the utility gains under these two optimal strategies. The following lemma summarizes the findings:

Lemma 1. *Consider a DM characterized by $q^{(0)}$. The optimal strategy of facing difficulties since the beginning (that is, whenever $\mu < q^{(0)}, q^d < q^e$) yields higher utility than the optimal strategy of facing them from an intermediate point in time (that is, whenever $q^{(0)} < \mu < q^d < q^e$).*

Also, in order to make the intermediate strategy (t) in C.2 and the always avoiding strategy (∞) in C.1 comparable, we consider, for the latter, the specific situation in which $q^{(0)} < q^d < q^e < \mu$. Notice that this is the only situation in C.1 in which $q^{(0)} < q^d$. Let $q^{(0)}$ be the same in both scenarios and focus on the case in which the only difference between C.1 and C.2 is that we increase μ, from C.2 to C.1, by decreasing K.[23] Since q^e and q^d do not depend on K, they remain unaltered. The following lemma summarizes the findings:

Lemma 2. *Consider a DM characterized by $q^{(0)}$. The optimal strategy of facing difficulties from an intermediate point in time (that is, whenever $q^{(0)} < \mu < q^d < q^e$) yields higher utility than the optimal strategy of always avoiding difficulties (that is, whenever $q^{(0)} < q^d < q^e < \mu$).*

With these two lemmas we conclude that optimal strategies involving the choice of difficulties at some point in time, yield higher utility than optimal strategies in which the DM always avoids difficulties.

3.4 Conclusions

Non-cognitive abilities have an impact in determining performance in dimensions of huge economic relevance, as labor market entry/ search decisions or educational attainments. We link, in a dynamic setting, non-cognitive abilities to the decision of when to deal with difficult but valuable tasks. We show how low disposition individuals always avoid difficulties and forego better economic opportunities while high disposition individuals are willing to deal with difficulties. The behavior of individuals that always avoid dealing with onerous tasks resembles procrastination results, without relying on the hyperbolic discounting assumption.[24] Also, individuals that get motivated by outcome achievements find optimal to jump into difficult tasks at some point in time.

[23] See the proof of Theorem 2, step 1.

[24] See Rubinstein (2003) for a discussion on this assumption.

3.5 Appendix. Proofs

Before proceeding we set some useful definitions. Let us denote by $u^{1,(0)}$ and $u^{2,(0)}$, the DM's long-run expected utility when she only experiences the deteriorated capacity state s_1, that is when $q = 0$, and when she only experiences the full capacity state s_2, that is when $q = 1$, respectively, under strategy (0). Similarly, let us denote by $u^{1,(1)}$ and $u^{2,(1)}$, the DM's long-run expected utility when she only experiences the deteriorated capacity state s_1, that is when $q = 0$, and when she only experiences the full capacity state s_2, that is when $q = 1$, respectively, under strategy (1).

Let us define functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ as $f(\lambda) = \lambda u^{2,(1)} + (1-\lambda)u^{1,(1)}$ and $g(\lambda) = \lambda u^{2,(0)} + (1-\lambda)u^{1,(0)}$, respectively. For $\lambda \in [0, 1]$, these functions are the convex combination of the DM's long-run expected utilities, when she experiences s_2 with probability $q = 1$ and $q = 0$, out of strategies (1) and (0), respectively. Furthermore, we say that a strategy (t) dominates strategy $(t + 1)$ whenever the long-run expected utility of (t) is higher than the one of $(t+1)$. Let $(t) > (t+1)$ denote this domination relationship. Recall that (t) denotes any strategy such that $0 < t < \infty$. We also say that strategy (0) dominates strategy (1) whenever the long-run expected utility of (0) is higher than the one of (1). Let $(0) > (1)$ denote this domination relationship.

Proof of Theorem 1. The proof is composed by two steps. In Step 1 we derive the threshold λ such $f(\lambda)$ and $g(\lambda)$ equalize. For such a λ, (1) and (0) yield the same long-run expected utility. For values higher or equal than λ then $(0) > (1)$. For values lower than λ then $(1) > (0)$. In step 2 we argue how this information is enough to set the optimal strategy, depending on the values of λ and q.

Step 1. If the DM experiences s_2 with probability $q = 1$, the long-run expected utility of strategy (0) is $u^{2,(0)} = Kp_{22} + \delta u^{2,(0)}$. If she experiences s_1 with probability $1 - q = 1$, the long-run expected utility of strategy (0) is $u^{1,(0)} = Kp_{12} + \delta u^{1,(0)}$. Similarly, when she experiences s_2 with probability $q = 1$, the long-run expected utility of strategy (1) is $u^{2,(1)} = p_{21} + \delta u^{2,(0)}$ whereas when she experiences s_1 with probability $1 - q = 1$, then $u^{1,(1)} = p_{11} + \delta u^{1,(0)}$. From previous definitions, $f(\lambda) = \lambda(u^{2,(1)} - u^{1,(1)}) + u^{1,(1)}$ and $g(\lambda) = \lambda(u^{2,(0)} - u^{1,(0)}) + u^{1,(0)}$. Solving $f(\lambda) = g(\lambda)$ for λ we get $\lambda((u^{2,(1)} - u^{1,(1)}) - (u^{2,(0)} - u^{1,(0)})) = u^{1,(0)} - u^{1,(1)}$. Notice that $u^{2,(0)} - u^{1,(0)} = K(p_{22} - p_{12})(1-\delta)^{-1}$ and $u^{1,(0)} = Kp_{12}(1-\delta)^{-1}$. Thus, $\lambda((p_{21} - p_{11}) - K(p_{22} - p_{12})) = Kp_{12} - p_{11}$ or $\lambda = (Kp_{12} - p_{11})((p_{21} - p_{11}) - K(p_{22} - p_{12}))^{-1}$. Assumption 1 implies that $p_{22} - p_{12} > p_{21} - p_{11}$, thus the denominator of the previous is different from zero, and λ exists. Also the denominator is negative and for values lower (respectively higher) than λ then $(1) > (0)$ (respectively $(0) > (1)$). For values equal to λ we assume that $(0) > (1)$ as well. Assumption 1 also implies that $p_{11}/p_{12} > p_{21}/p_{22}$.[25] Thus,

[25] By assumption 1, $p_{11} - p_{12} > p_{21} - p_{22}$. It implies that $(p_{11} - p_{12})(p_{12}) - 1 > (p_{21} - p_{22})(p_{22})^{-1}$ or equivalently $p_{11}/p_{12} > p_{21}/p_{22}$.

$\partial\lambda/\partial K = \dfrac{p_{21}p_{12} - p_{22}p_{11}}{((p_{21} - p_{11}) - K(p_{22} - p_{12}))^2} < 0$. Notice than when $K = p_{21}/p_{22}$ then $\lambda = 1$ and when $K = p_{11}/p_{12}$ then $\lambda = 0$. It then follows that when $K < p_{21}/p_{22}$ then $\lambda > 1$, and no matter q, $(1) > (0)$. On the contrary, when $K > p_{11}/p_{12}$ then $\lambda < 0$, and no matter q, $(0) > (1)$. We focus on the interesting case such that $p_{11}/p_{12} < K < p_{11}/p_{12}$ and $\lambda \in (0, 1)$. We thus conclude that when $q \geq \lambda$ then $(0) > (1)$ and when $q < \lambda$ then $(1) > (0)$.

Step 2. We set here the optimal strategies. Two cases arise depending on the relation between q and λ:

C.1. Suppose that $q \geq \lambda$. By step 1, $(0) > (1)$. Let us compare any pair of intermediate strategies (t) and $(t+1)$. We have that $u((t)) = \sum_{i=0}^{t-1} \delta^i (qp_{21} + (1-q)p_{11}) + K \sum_{i=t}^{\infty} \delta^i (qp_{22} + (1-q)p_{12})$ and $u((t+1)) = \sum_{i=0}^{t-1} \delta^i (qp_{21} + (1-q)p_{11}) + \delta^t (qp_{21} + (1-q)p_{11}) + K \sum_{i=t+1}^{\infty} \delta^i (qp_{22} + (1-q)p_{12})$. Notice that up to the point in time $t-1$, (t) and $(t+1)$ yield the same utility. Notice also that from time t on, the comparison is between (0) and (1), evaluated from the point of view of time t. Since it is always the case that $q \geq \lambda$, it follows that $(0) > (1)$, from the point of view of time t. That is so because we can consider the process as starting at time t and thus, apply step 1. As a consequence, for any pair of intermediate strategies, (t) and $(t+1)$, it follows that $(t) > (t+1)$. It is useful to observe that $\lim_{i \longrightarrow \infty} u((t+i)) = u((\infty))$. We then conclude that $(0) > (1) > ... > (t) > (t+1) > ... > (\infty)$ holds. In this case (0) is optimal.

C.2. Suppose that $q < \lambda$. To conclude that $(0) < (1) < ... < (t) < (t+1) < ... < (\infty)$ we use a similar reasoning as above and thus omit it here. In this case (∞) is optimal.

■

Proof of the Remark. See the proof of Theorem 1 for the expression of λ and its relation with K. We analyze here how λ varies with the probabilities of success. Let $x \equiv (p_{21} - p_{11} - K(p_{22} - p_{12}))^2$ be the denominator in the following derivatives. We have that: $\partial\lambda/\partial p_{11} = (Kp_{22} - p_{21})x^{-1}$, $\partial\lambda/\partial p_{12} = K(p_{21} - Kp_{22})x^{-1}$, $\partial\lambda/\partial p_{21} = (p_{11} - Kp_{12})x^{-1}$ and $\partial\lambda/\partial p_{22} = K(Kp_{12} - p_{11})x^{-1}$. Since $K \in (p_{21}/p_{22}, p_{11}/p_{12})$ then $\partial\lambda/\partial p_{i1} > 0$ and $\partial\lambda/\partial p_{i2} < 0$ with $i = 1, 2$.

■

Proof of Proposition 1. Recall that $u((0)) = K \sum_{i=0}^{\infty} \delta^i (qp_{22} + (1-q)p_{12}) = (K(qp_{22} + (1-q)p_{12})(1-\delta)^{-1}$ and $u((\infty)) = \sum_{i=0}^{\infty} \delta^i (qp_{21} + (1-q)p_{11}) = (qp_{21} + (1-q)p_{11})(1-\delta)^{-1}$. Since $\partial u((0))/\partial q = K(p_{22} - p_{12})(1-\delta)^{-1} > 0$ and $\partial u((\infty))/\partial q = (p_{21} - p_{11})(1-\delta)^{-1} > 0$, utilities are increasing and linear in q. Finally, assumption 1 implies that $p_{22} - p_{12} > p_{21} - p_{11}$, then the marginal return of an increase in q is the highest under (0). ■

Before the proof of Theorem 2 let us set two useful claims:

Claim 1. *Consider that the DM repeatedly faces easy tasks. Then, at every time t, $q^{(0)} > q^{(t)} > q^{(t+1)} > q^e$ whenever $q^{(0)} > q^e$ and $q^{(0)} < q^{(t)} < q^{(t+1)} < q^e$ whenever $q^{(0)} < q^e$.*

Proof of Claim 1. The proof is by induction. Let us focus on the case in which $q^{(0)} > q^e$. We first prove that for $t = 1$, $q^{(0)} > q^{(1)} > q^e$ holds. We set the induction part afterwards.

Step 1. $q^{(0)} > q^{(1)} > q^e$. In showing that $q^{(0)} > q^{(1)}$, we compare the initial probability of experiencing s_2, with its first perturbation, after having decided to face an easy task. Consider expression (3.1) in the main body:

$$[q^{(0)} \quad 1-q^{(0)}]\begin{bmatrix} p_{21} & 1-p_{21} \\ p_{11} & 1-p_{11} \end{bmatrix} = [q^{(1)} \quad 1-q^{(1)}].$$

We have that $q^{(1)} = q^{(0)}p_{21} + (1-q^{(0)})p_{11}$. Recall that $q^{(0)} > q^e$ and $q^e = p_{11}(1 - p_{21} + p_{11})^{-1}$. Thus, $q^{(0)} > p_{11}(1-p_{21}+p_{11})^{-1}$ or equivalently $q^{(0)}(1-p_{21}+p_{11}) > p_{11}$. We rewrite this expression as $q^{(0)} > q^{(0)}p_{21} + (1-q^{(0)})p_{11}$. The RHS of this expression is exactly $q^{(1)}$. In showing that $q^{(1)} > q^e$ we proceed by contradiction. Suppose that $q^{(1)} < q^e$ holds, that is, $q^{(0)}p_{21} + (1-q^{(0)})p_{11} < q^e$. This is equivalent to $q^{(0)}p_{21} + (1 - q^{(0)})p_{11} < p_{11}(1-p_{21}+p_{11})^{-1}$ or $(p_{11}+q^{(0)}(p_{21}-p_{11}))(1-p_{21}+p_{11}) < p_{11}$. Rearranging terms it becomes $p_{11} - p_{11}(p_{21} - p_{11}) + q^{(0)}(p_{21} - p_{11})(1 - p_{21} + p_{11}) < p_{11}$. This is equivalent to $q^{(0)} < p_{11}(1 - p_{21} + p_{11})^{-1} = q^e$, contradicting our initial assumption. Thus, $q^{(0)} > q^{(1)} > q^e$ holds.

Step 2. If for an arbitrary t, $q^{(t)} > q^{(t+1)} > q^e$ holds, for $q^{(t+1)}$ we have:

$$\left[q^{(t+1)} \quad 1-q^{(t+1)}\right]\begin{bmatrix} p_{21} & 1-p_{21} \\ p_{11} & 1-p_{11} \end{bmatrix} = [q^{(t+2)} \quad 1-q^{(t+2)}].$$

In concluding that $q^{(t+1)} > q^{(t+2)} > q^e$ we use exactly the same reasoning than in the previous step. We conclude that $q^{(0)} > q^{(t)} > q^{(t+1)} > ... > q^e$ holds. The case in which $q^{(0)} < q^e$ relies on the same argument. The same analysis goes through for describing the relation between $q^{(t)}$ and q^d. We thus omit the proofs here. ■

Claim 2. $q^{(i)} = q^{(0)}(T^k)^i + p_{1k}\sum_{j=0}^{i-1}(T^k)^j$ *and* $q^{(t+i)} = q^{(t)}(T^k)^i + p_{1k}\sum_{j=0}^{i-1}(T^k)^j$ *with* $k = e, d$.

Proof of Claim 2. Recall that $T^d = p_{22} - p_{12}$ and $T^e = p_{21} - p_{11}$. By expression (3.1) in the main body, $q^1 = q^{(0)}T^k + p_{1k}$. Also $q^2 = q^{(1)}T^k + p_{1k} = (q^{(0)}T^k + p_{1k})T^k + p_{1k} = q^{(0)}(T^k)^2 + p_{1k}T^k + p_{1k}$. In general $q^{(i)} = q^{(0)}(T^k)^i + p_{1k}(T^k)^{i-1} + ... + p_{1k}T^k + p_{1k}$ or $q^{(i)} = q^{(0)}(T^k)^i + p_{11}\sum_{j=0}^{t-1}(T^k)^j$. To conclude that $q^{(t+i)} = q^{(t)}(T^k)^i + p_{1k}\sum_{j=0}^{i-1}(T^k)^j$ we follow a similar reasoning. We thus omit it here. ■

Proof of Theorem 2. We follow the same steps than in the proof of Theorem 1.

Step 1. Consider expression (3.1) in the main body. When the DM experiences s_2 with probability $q^{(0)} = 1$, the long-run expected utility out of strategy (0) is $u^{2,(0)} = Kp_{22} + \delta(p_{22}u^{2,(0)} + (1-p_{22})u^{1,(0)})$. When she experiences s_1 with probability $1 - q^{(0)} = 1$, the long-run expected utility of strategy (0) is $u^{1,(0)} = Kp_{12} + \delta(p_{12}u^{2,(0)} + (1-p_{12})u^{1,(0)})$. Solving for $u^{1,(0)}$ and $u^{2,(0)}$ we get $u^{2,(0)} = \dfrac{K(p_{22} - \delta T^d)}{(1-\delta)(1-\delta T^d)}$ and that $u^{1,(0)} = \dfrac{Kp_{12}}{(1-\delta)(1-\delta T^d)}$. Thus, $u^{2,(0)} - u^{1,(0)} = \dfrac{KT^d}{1-\delta T^d}$. When she experiences s_2 with probability $q^{(0)} = 1$, the long-run expected utility out of strategy (1) is $u^{2,(1)} = p_{21} + \delta(p_{21}u^{2,(0)} + (1-p_{21})u^{1,(0)})$. Similarly, when she experiences s_1, with probability $1 - q^{(0)} = 1$, the long-run expected utility out of strategy (1) can is $u^{1,(1)} = p_{11} + \delta(p_{11}u^{2,(0)} + (1-p_{11})u^{1,(0)})$. Now, $f(\mu) = \mu(u^{2,(1)} - u^{1,(1)}) + u^{1,(1)}$ and $g(\mu) = \mu(u^{2,(0)} - u^{1,(0)}) + u^{1,(0)}$ Solving for μ such that $f(\mu) = g(\mu)$ we get $\mu = \dfrac{u^{1,(0)} - u^{1,(1)}}{(u^{2,(1)} - u^{1,(1)}) - (u^{2,(0)} - u^{1,(0)})}$. With respect to the numerator, $u^{1,(0)} - u^{1,(1)} = (1-\delta)u^{1,(0)} - p_{11} - \delta p_{11}(u^{2,(0)} - u^{1,(0)})$ or equivalently $u^{1,(0)} - u^{1,(1)} = \dfrac{(1-\delta)Kp_{12}}{(1-\delta)(1-\delta T^d)} - p_{11} - \dfrac{\delta p_{11} KT^d}{1-\delta T^d} = \dfrac{K(p_{12} - \delta p_{11}T^d) - (p_{11} - \delta p_{11}T^d)}{1-\delta T^d}$. Regarding the denominator, $u^{2,(1)} - u^{1,(1)} = T^e + \delta T^e(u^{2,(0)} - u^{1,(0)})$ and $u^{2,(1)} - u^{1,(1)} - (u^{2,(0)} - u^{1,(0)}) = T^e - (1-\delta T^e)(u^{2,(0)} - u^{1,(0)})$. This is equivalent to $\dfrac{T^e(1-\delta T^d) - KT^d(1-\delta T^e)}{(1-\delta T^d)}$. Thus, $\mu = \dfrac{K(p_{12} - \delta p_{11}T^d) - (p_{11} - \delta p_{11}T^d)}{T^e(1-\delta T^d) - KT^d(1-\delta T^e)}$.[26] Since by assumption 1, $T^d > T^e$ the denominator is different from zero, hence μ is a real number. Since by assumption 1, the denominator is negative, for values lower than μ then (1) > (0) and for values higher than it, (0) > (1). For values equal to μ we assume that (0) > (1) as well. Also by assumption 1, $\partial\mu/\partial K = \dfrac{(1-\delta T^d)(p_{21}p_{12} - p_{11}p_{22})}{(T^e(1-\delta T^d) - KT^d(1-\delta T^e))^2} < 0$. Furthermore, when $K = \dfrac{p_{11} - \delta p_{11}T^d}{p_{12} - \delta p_{11}T^d}$ then $\mu = 0$ and when $K = \dfrac{p_{21} - \delta p_{21}T^d}{p_{22} - \delta p_{21}T^d}$ then $\mu = 1$. Thus, it has to be that $\dfrac{p_{21} - \delta p_{21}T^d}{p_{22} - \delta p_{21}T^d} < \dfrac{p_{11} - \delta p_{11}T^d}{p_{12} - \delta p_{11}T^d}$. It also has to be that when $K > \dfrac{p_{11} - \delta p_{11}T^d}{p_{12} - \delta p_{11}T^d}$ then $\mu < 0$. In this case no matter $q^{(0)}$, (0) > (1). In contrast, when $K < \dfrac{p_{21} - \delta p_{21}T^d}{p_{22} - \delta p_{21}T^d}$, then $\mu > 1$ and no matter $q^{(0)}$, (1) > (0). The interesting case is such that $K \in \left(\dfrac{p_{21} - \delta p_{21}T^d}{p_{22} - \delta p_{21}T^d}, \dfrac{p_{11} - \delta p_{11}T^d}{p_{12} - p_{11}\delta T^d}\right)$ and $\mu \in (0, 1)$. We conclude that when $q^{(0)} \geq \mu$ then (0) > (1) and when $q^{(0)} < \mu$ then (1) > (0).

[26] When the DM does not care about the future, that is, when $\delta = 0$, $\lambda = \mu$. That q does not vary over time is conceptually equivalent to think about a DM making one period decisions without consequences on her subsequent states. Additionally, $\partial\mu/\partial\delta = K(K-1)(p_{11}p_{22} - p_{21}p_{22}) > 0$, meaning that the more the DM cares about the future the more she postpones difficult tasks, where good outcomes are less frequent.

Step 2. We set here the optimal strategies. Consider first, that $\mu \leq q^{(0)}$. By step 1, $(0) > (1)$ from the point of view of $q^{(0)}$. Three cases arise:

C.1. Suppose that $\mu \leq q^d < q^e, q^{(0)}$. Let us compare any pair of intermediate strategies, (t) and $(t+1)$. We thus evaluate $u((t)) = \sum_{i=0}^{t-1}(q^{(i)}p_{21} + (1-q^{(i)})p_{11}) + K\sum_{i=t}^{\infty}\delta^i(q^{(i)}p_{22} + (1-q^{(i)})p_{12})$ versus $u((t+1)) = \sum_{i=0}^{t-1}(q^{(i)}p_{21} + (1-q^{(i)})p_{11}) + \delta^t(q^{(t)}p_{21} + (1-q^{(t)})p_{11}) + K\sum_{i=t+1}^{\infty}\delta^i(q^{(i)}p_{22} + (1-q^{(i)})p_{12})$. Notice that up to time $t-1$, both expressions yield the same utility. From time t on, the comparison is between (0) and (1), from the point of view of $q^{(t)}$. Notice that for any strategy (t), $q^{(t)}$ results from have been dealing with easy tasks up to time $t-1$. Thus, by Claim 1, $q^{(t)} > \mu$. This implies that, from the point of view of $q^{(t)}$, $(0) > (1)$. That is so because we can consider the process as starting at time t, and thus apply step 1. As a consequence, for any pair of intermediate strategies (t) and $(t+1)$, it follows that $(t) > (t+1)$. Recall that $\lim_{i \longrightarrow \infty} u((t+i)) = u((\infty))$.[27] We thus establish that $(0) > (1)... > (t) > (t+1) > ... > (\infty)$. Then (0) is optimal.

Assumption 2 does not play any role in C.1, that is, the DM's optimal strategy is within the class of strategies that it prescribes. However, it does in C.2 and C.3 below. In both cases the DM would have found optimal to start with difficult tasks and to switch to easy ones at some point in time. We look for the optimal strategies within the ones prescribed by assumption 2.

C.2. Suppose that $q^d < \mu \leq q^e, q^{(0)}$. Let us compare (t) and $(t+1)$, as above. We then evaluate $u((t))$ and $u((t+1))$ as defined in $C.1$. The relevant comparison is between (0) and (1) from the point of view of $q^{(t)}$. Notice that $q^{(t)}$ results from dealing with easy tasks up to time $t-1$. Thus, by Claim 1, $q^{(t)} \geq \mu$. Then, by step 1, from the point of view of $q^{(t)}$, $(0) > (1)$ and, as a consequence, $(t) > (t+1)$. We thus establish that $(0) > (1) > ... > (t) > (t+1) > ... > (\infty)$, being (0) optimal.

C.3. Suppose that $q^d < q^e < \mu \leq q^{(0)}$. Let us compare (t) and $(t+1)$ as above. We then evaluate $u((t))$ and $u((t+1))$ as defined in C.1. The relevant comparison is between (0) and (1) from the point of view of $q^{(t)}$. Suppose that strategy (t) is such that $q^{(t)} > \mu$. Again, from the point of view of $q^{(t)}$, $(0) > (1)$. As a consequence, $(t) > (t+1)$. Suppose that strategy (t^*-1) is such that $q^{(t^*-1)} = \mu$.[28] Notice that $t < t^* - 1$, by Claim 1 and, at least, $t^* - 1 = t + 1$. Thus, from the point of view of $q^{(t^*-1)}$, $(0) > (1)$. As a consequence, $(t^*-1) > (t^*)$. Suppose that strategy (t) is such that $q^{(t)} < \mu$ for the first time. By Claim 1, this point in time has to be exactly t^*. Thus, from the point of view of $q^{(t^*)}$, $(1) > (0)$. As a consequence $(t^*+1) > (t^*)$. Suppose that strategy (t) is any is such that $q^{(t)} < \mu$. Notice that $t^* < t$, by Claim 1 and, at least, $t = t^* + 1$. Thus, from the point of view of $q^{(t)}$, $(1) > (0)$. As a consequence, $(t) < (t+1)$, in particular, $(t^*+1) < (t^*+2)$. In general we have that

[27] This observation applies to the remaining cases. We omit it in what follows.

[28] The same analysis follows if we consider that $q^{(t^*-1)} < \mu$.

$(0) > (1) > ... > (t^* - 1) > (t^*) < (t^* + 1) < (t^* + 2) < ... < (\infty)$. An intermediate strategy (t) is then the least preferred and the optimal is either (0) or (∞).

Consider now that $q^{(0)} < \mu$. By step 1, $(1) > (0)$ from the point of view of $q^{(0)}$. Three cases arise. As we use parallel arguments than above, we go briefly over them. We compare $u((t))$ and $u((t + 1))$ in every case. The relevant comparison will be between (0) and (1), from the point of view of $q^{(t)}$:

C.1. Suppose that $q^{(0)}, q^d < q^e \leq \mu$. Notice that for any strategy (t), $q^{(t)}$ results from have been dealing with easy tasks up to time $t - 1$. Thus, by Claim 1, $q^{(t)} < \mu$. Thus by step 1, $(1) > (0)$ from the point of view of $q^{(t)}$. As a consequence, for any pair of strategies, $(t + 1) > (t)$. Thus, $(0) < (1) < ... < (t) < (t + 1) < ... < (\infty)$ and the optimal strategy is (∞).

C.2. Suppose that $q^{(0)} < \mu \leq q^d < q^e$. Suppose that strategy (t) is such that $q^{(t)} < \mu$. Thus, from the point of view of $q^{(t)}$, $(1) > (0)$. As a consequence, $(t+1) > (t)$. Suppose that strategy (t^*) is such that $q^{(t^*)} = \mu$. Notice that $t < t^*$, by Claim 1 and, at least, $t^* = t + 1$. Thus, from the point of view of $q^{(t^*)}$, $(0) > (1)$. As a consequence, $(t^*) > (t^* + 1)$. Suppose that strategy (t) is such that $q^{(t)} > \mu$ for the first time. By Claim 1, this point in time has to be exactly $t^* + 1$. Thus, from the point of view of $q^{(t^*+1)}$, $(0) > (1)$. As a consequence $(t^*+1) > (t^*+2)$. Suppose that strategy (t) is any other such that $q^{(t)} > \mu$. Notice that $t^* + 1 < t$, by Claim 1 and, at least, $t = t^* + 2$. Thus, from the point of view of $q^{(t)}$, $(0) > (1)$. As a consequence $(t) > (t + 1)$. We thus have that, $(0) < (1) < ... < (t^* - 1) < (t^*) > (t^* + 1) > (t^* + 2) > ... > (\infty)$. Then, (t) is optimal.

Assumption 2 does not play a role in $C.1$ and $C.2$. However, it does in the last case. In it, the DM would have preferred to switch from difficult to easy tasks at some point. We look for the optimal strategies within the ones prescribed by assumption 2.

C.3. Suppose that $q^{(0)}, q^d \leq \mu < q^e$. Suppose that strategy (t) is such that $q^{(t)} < \mu$. Thus, from the point of view of $q^{(t)}$, $(1) > (0)$. As a consequence, $(t+1) > (t)$. Suppose that strategy (t^*) is such that $q^{(t^*)} = \mu$. Notice that $t < t^*$, by Claim 1 and, at least, $t^* = t + 1$. Thus, from the point of view of $q^{(t^*)}$, $(0) > (1)$. As a consequence, $(t^*) > (t^* + 1)$. Suppose that strategy (t) is such that $q^{(t)} > \mu$ for the first time. This point in time is exactly $t^* + 1$. Thus, from the point of view of $q^{(t^*+1)}$, $(0) > (1)$. As a consequence $(t^* + 1) > (t^* + 2)$. Suppose that strategy (t) is any strategy such that $q^{(t)} > \mu$. Notice that $t^* + 1 < t$, by Claim 1 and, at least, $t = t^* + 2$. Thus, from the point of view of $q^{(t)}$, $(0) > (1)$. As a consequence, $(t) > (t + 1)$. Summing up we have that $(0) < (1) < ... < (t^* - 1) < (t^*) > (t^* + 1) > (t^* + 2) > ... > (\infty)$. Then, (t) is optimal.

■

Proof of Proposition 2. We have three cases depending on the optimal strategy:

C.1.(0) is optimal. We have that $u((0)) = K \sum_{i=0}^{\infty} \delta^i (q^{(i)} p_{22} + (1 - q^{(i)}) p_{12})$.

By Claim 2, $q^{(i)} = q^{(0)}(T^d)^i + p_{12}\sum_{j=0}^{i-1}(T^d)^j$. Plugging $q^{(i)}$ in the previous expression we have that $u((0)) = K\sum_{i=0}^{\infty}\delta^i(q^{(i)}T^d + p_{12})$ or $KT^dq^{(0)}\sum_{i=0}^{\infty}(\delta T^d)^i + p_{12}T^dK\sum_{i=0}^{\infty}\delta^i\sum_{j=0}^{i-1}(T^d)^j + Kp_{12}\sum_{i=0}^{\infty}\delta^i$. Then $\partial u((0))/\partial q^{(0)} = KT^d(1-\delta T^d)^{-1} > 0$.

C.2. (∞) is optimal. We have that $u((\infty)) = \sum_{i=0}^{\infty}\delta^i(q^{(i)}p_{21} + (1-q^{(i)})p_{11})$. We follow exactly the same reasoning than in C.1, and thus omit it here. In this case $\partial u((\infty))/\partial q^{(0)} = T^e(1-\delta T^e)^{-1} > 0$.

In both cases utility is increasing and linear in $q^{(0)}$. Since $T^d > T^e$, the marginal return of an increase in $q^{(0)}$ is higher under (0) than under (∞).

C.3. (t) is optimal. We have that $u((t)) = \sum_{i=0}^{t-1}\delta^i(q^{(i)}p_{21} + (1-q^{(i)})p_{11}) + K\sum_{i=t}^{\infty}\delta^i(q^{(i)}p_{22} + (1-q^{(i)})p_{12})$. Let us focus first on the first part of the expression, that is, $\sum_{i=0}^{t-1}\delta^i(q^{(i)}p_{21} + (1-q^{(i)})p_{11})$. By Claim 2, $q^{(i)} = q^{(0)}(T^e)^i + p_{11}\sum_{j=0}^{i-1}(T^e)^j$. Thus, $\sum_{i=0}^{t-1}\delta^i(q^{(i)}T^e + p_{11}) = \sum_{i=0}^{t-1}\delta^i((q^{(0)}(T^e)^i + p_{11}\sum_{j=0}^{i-1}(T^e)^j)T^e + p_{11})$. This expression is equivalent to $q^{(0)}T^e\sum_{i=0}^{t-1}\delta^i(T^e)^i + T^e\sum_{i=0}^{t-1}\delta^ip_{11}\sum_{j=0}^{i-1}(T^e)^j + \sum_{i=0}^{t-1}\delta^ip_{11}$. Its derivative with respect to $q^{(0)}$ is $T^e\sum_{i=0}^{t-1}(\delta T^e)^i > 0$. Consider now the second part, that is, $K\sum_{i=t}^{\infty}\delta^i(q^{(i)}p_{22} + (1-q^{(i)})p_{12})$. By Claim 2, $q^{(t+i)} = q^{(t)}(T^d)^i + p_{12}\sum_{j=0}^{i-1}(T^d)^j$. Thus, we rewrite the previous expression as $K(\sum_{i=0}^{\infty}\delta^{t+i}((q^{(t)}(T^d)^i + p_{12}\sum_{j=0}^{i-1}(T^d)^j)T^d + p_{12}))$. This is equivalently rewritten as $K(q^{(t)}T^d\sum_{i=0}^{\infty}\delta^{t+i}(T^d)^i + T^d\sum_{i=0}^{\infty}\delta^{t+i}p_{12}\sum_{j=0}^{\infty}(T^d)^j + \sum_{i=0}^{\infty}\delta^{t+i}p_{12}))$. By Claim 2 we express the part depending on $q^{(t)}$ as $(q^{(0)}(T^d)^t + p_{12}\sum_{i=0}^{t-1}(T^d)^i)K\delta^tT^d(1-\delta T^d)^{-1}$. Taking derivatives w.r.t $q^{(0)}$ we get $K\delta^t(T^d)^{t+1}(1-\delta T^d)^{-1} > 0$. Summing up, we have that $\partial u((t))/\partial q^{(0)} = T^e\sum_{i=0}^{t-1}(\delta T^e)^i + K\delta^t(T^d)^{t+1})(1-\delta T^d)^{-1} > 0$.

We now compare the return of a marginal increase in $q^{(0)}$, in the aforementioned strategies. Notice that $u((0)) = K(\sum_{i=0}^{t-1}\delta^i(q^{(i)}T^d + p_{12}) + \sum_{i=t}^{\infty}\delta^i(q^{(i)}T^d + p_{12}))$ and $u((\infty)) = \sum_{i=0}^{t-1}\delta^i(q^{(i)}T^e + p_{11}) + \sum_{i=t}^{\infty}\delta^i(q^{(i)}T^e + p_{11})$. We use similar algebra as above to conclude that $\partial u((0))/\partial q^{(0)} = KT^d\sum_{i=0}^{t-1}(\delta T^d)^i + K\delta^t(T^d)^{t+1}(1-\delta T^d)^{-1}$ and $\partial u((\infty))/\partial q^{(0)} = T^e\sum_{i=0}^{t-1}(\delta T^e)^i + \delta^t(T^e)^{t+1}(1-\delta T^e)^{-1}$ Since $T^d > T^e$, we have that $\partial u((0))/\partial q^{(0)} > \partial u((t))/\partial q^{(0)}$ and $\partial u((t))/\partial q^{(0)} > \partial u((\infty))/\partial q^{(0)}$. We also compare the marginal return of any pair of intermediate strategies $(t-1)$ and (t). In this case $t-1 \geq 1$. We have that $\partial u((t-1))/\partial q^{(0)} = T^e\sum_{i=0}^{t-2}(\delta T^e)^i + K\delta^{t-1}(T^d)^t(1-\delta T^d)^{-1}$ and $\partial u((t))/\partial q^{(0)} = T^e\sum_{i=0}^{t-1}(\delta T^e)^i + K\delta^t(T^d)^{t+1}(1-\delta T^d)^{-1}$. The latter expression minus the former yields $\delta^{t-1}(K(T^d)^t - (T^e)^t)$, which is positive since $T^d > T^e$. ■

Proof of Lemma 1. Consider C.2 in table 2 in the main body. Let us denote by $q^{(0)'}$ the initial probability in any of the cases in which (0) is optimal. Let us also denote by $q^{(0)}$ the initial probability in the case in which (t) is optimal. Notice that $q^{(0)'} > q^{(0)}$. Consider the utility of (t) when the DM is characterized by $q^{(0)'}$, that is, when (0) is optimal. Notice that the utility of (t) is higher when the DM is characterized by $q^{(0)'}$ than when she is characterized by $q^{(0)}$ and precisely (t) is optimal. To see this, notice that up to $t-1$ the DM faces easy tasks. Since $q^{(0)'} > q^{(0)}$, by Claim 2, $q^{(i)'} > q^{(i)}$ at every $i \leq t-1$. Consider now points in time $i \geq t$. By Claim 1, $q^{(i)}$ approaches q^d from below without exceeding it. Also, $q^{(i)'}$ may approach q^d from below or above, without exceeding it.[29] When $q^{(i)'}$ approaches q^d from below, by Claim 2, $q^{(i)'} > q^{(i)}$. When $q^{(i)'}$ approaches q^d from above by Claim 1, $q^{(i)'} > q^d > q^{(i)}$. Since current expected utility at every time t, that is, $K(q^{(t)}p_{22} + (1 - q^{(t)}p_{12})$, is increasing $q^{(t)}$, it has to be that $u((t))$ is higher under $q^{(0)'}$ than under $q^{(0)}$. Also, under $q^{(0)'}$, $u((0)) > u((t))$ by optimality. We thus conclude that the optimal strategy (0) yields higher utility than the optimal strategy (t).

■

Proof of Lemma 2. Consider that the DM is characterized by $q^{(0)}$. By the proof of Theorem 2, under $q^{(0)} < q^d < q^e < \mu'$, (∞) is optimal whereas under $q^{(0)} < \mu < q^d < q^e$, (t) is optimal. Recall that μ is decreasing in K, thus $\mu < \mu'$ is associated to $K > K'$. By optimality of (t) we have that $\sum_{i=0}^{t-1} \delta^i(q^{(i)}T^e + p_{11}) + K \sum_{i=t}^{\infty} \delta^i(q^{(i)}T^d + p_{12}) > \sum_{i=0}^{t-1} \delta^i(q^{(i)}T^e + p_{11}) + \sum_{i=t}^{\infty} \delta^i(q^{(i)}T^e + p_{11})$. Since we consider the case in which $q^{(0)}$ as well as probabilities of success affecting q^d and q^e are the same, the RHS of this expression brings exactly the same utility that when $q^{(0)} < q^d < q^e < \mu'$, and hence (∞) is optimal. Thus, $u((t))$ under $q^{(0)} < \mu < q^d < q^e$ is higher than $u((\infty))$ under $q^{(0)} < q^d < q^e < \mu'$.

■

[29] For Claim 1 to apply we consider, as in previous proofs, the process as starting at time $i = t$. Also, the behavior of $q^{(i)'}$ depends on whether in approaching q^e, q^d is exceeded or not.

Bibliography

Daron Acemoglu and Asuman Ozdaglar. Opinion dynamics and learning in social networks. *Dynamic Games and Applications*, 1(1):3–49, 2011.

Daron Acemoglu, Asuman Ozdaglar, and Ali ParandehGheibi. Spread of (mis) information in social networks. *Games and Economic Behavior*, 70(2):194–227, 2010.

Daron Acemoglu, Giacomo Como, Fabio Fagnani, and Asuman Ozdaglar. Opinion fluctuations and disagreement in social networks. *Mathematics of Operations Research*, 38(1):1–27, 2013.

Alberto Alesina and Guido Tabellini. A positive theory of fiscal deficits and government debt. *The Review of Economic Studies*, 57(3):403–414, 1990.

Nageeb Ali. Learning self-control. *The Quarterly Journal of Economics*, 126(2):857–893, 2011.

James Andreoni and Tymofiy Mylovanov. Diverging opinions. *American Economic Journal: Microeconomics*, pages 209–232, 2012.

Tom M. Apostol. *Análisis matemático*. Reverté, 1977.

Hal R. Arkes and Catherine Blumer. The psychology of sunk cost. *Organizational behavior and human decision processes*, 35(1):124–140, 1985.

John William Atkinson and Norman T. Feather. *A theory of achievement motivation*, volume 66. Wiley New York, 1966.

Pau Balart and Antonio Cabrales. La maratón de pisa: La perseverancia como factor del éxito en una prueba de competencias. 2014.

Abhijit Banerjee and Drew Fudenberg. Word-of-mouth learning. *Games and Economic Behavior*, 46(1):1–22, 2004.

Roland Bénabou and Jean Tirole. Self-confidence and personal motivation. *The Quarterly Journal of Economics*, 117(3):871–915, 2002.

Douglas Bernheim and Antonio Rangel. Addiction and cue-triggered decision processes. *American Economic Review*, 94(5):1558–1590, 2004.

Marianne Bertrand. New perspectives on gender. *Handbook of labor economics*, 4: 1543–1590, 2011.

Pedro Bordalo, Nicola Gennaioli, and Andrei Shleifer. Salience and consumer choice. *Journal of Political Economy*, 121(5):803–843, 2013.

Jordi Brandts, Ayça Ebru Giritligil, and Roberto A. Weber. An experimental study of persuasion bias and social influence in networks. Technical report, BELIS, Istanbul Bilgi University, 2014.

Tiago Cavalcanti, Chryssi Giannitsarou, and Charles Johnson. Network cohesion. *Mimeograph, University of Cambridge*, 2012.

Arun Chandrasekhar, Horacio Larreguy, and Juan Pablo Xandri. Testing models of social learning on networks: Evidence from a framed field experiment. *Work. Pap., Mass. Inst. Technol., Cambridge, MA*, 2012.

Samprit Chatterjee and Eugene Seneta. Towards consensus: some convergence theorems on repeated averaging. *Journal of Applied Probability*, pages 89–97, 1977.

Hsiao-Chi Chen, James W. Friedman, and Jacques-Francois Thisse. Boundedly rational nash equilibrium: a probabilistic choice approach. *Games and Economic Behavior*, 18(1):32–54, 1997.

Andrew E. Clark. Job satisfaction and gender: why are women so happy at work? *Labour economics*, 4(4):341–372, 1997.

Luca Corazzini, Filippo Pavesi, Beatrice Petrovich, and Luca Stanca. Influential listeners: An experiment on persuasion bias in social networks. *European Economic Review*, 56(6):1276–1288, 2012.

Patricio S. Dalton, Sayantan Ghosal, and Anandi Mani. Poverty and aspirations failure. *The Economic Journal*, 2014.

Morris H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.

Peter M. DeMarzo, Jeffrey Zwiebel, and Dimitri Vayanos. Persuasion bias, social influence, and unidimensional opinions. *The Quarterly Journal of Economics*, 118 (3):909–968, 2003.

Federico Echenique and Roland G. Fryer. A measure of segregation based on social interactions. *The Quarterly Journal of Economics*, 122(2):441–485, 2007.

Glenn Ellison and Drew Fudenberg. Rules of thumb for social learning. *Journal of political Economy*, 101(4):612–643, 1993.

Noah E. Friedkin and Eugene C. Johnsen. Social influence networks and opinion change. *Advances in group processes*, 16(1):1–29, 1999.

Douglas Gale and Shachar Kariv. Bayesian learning in social networks. *Games and Economic Behavior*, 45(2):329–346, 2003.

Benjamin Golub and Matthew O. Jackson. Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2:112–149, 2010.

Benjamin Golub and Matthew O. Jackson. How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics*, 127(3):1287–1338, 2012.

Cleotilde Gonzalez. Task workload and cognitive abilities in dynamic decision making. *Human factors: The Journal of the Human factors and Ergonomics Society*, 47(1): 92–101, 2005.

Veronika Grimm and Friederike Mengel. An experiment on belief formation in networks. *Available at SSRN 2361007*, 2014.

James J. Heckman, Jora Stixrud, and Sergio Urzua. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3):411–482, 2006.

Rainer Hegselmann and Ulrich Krause. Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3), 2002.

Ilse C. F. Ipsen and Teresa M. Selee. Ergodicity coefficients defined by vector norms. *SIAM Journal on Matrix Analysis and Applications*, 32(1):153–200, 2011.

Matthew O. Jackson. *Social and economic networks*. Princeton University Press, 2008.

Oliver P. John and Sanjay Srivastava. The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, 2(1999):102–138, 1999.

Gueorgi Kossinets and Duncan J. Watts. Empirical analysis of an evolving social network. *Science*, 311(5757):88–90, 2006.

Botond Kőszegi and Adam Szeidl. A model of focusing in economic choice. *The Quarterly Journal of Economics*, 128(1):53–104, 2013.

Ulrich Krause. A discrete nonlinear and non-autonomous model of consensus formation. *Communications in difference equations*, pages 227–236, 2000.

Alle Leizarowitz. On infinite products of stochastic matrices. *Linear algebra and its applications*, 168:189–219, 1992.

Elliot Liebow. *Tally's corner: A study of Negro streetcorner men.* Boston: Little Brown, 1967.

Miller McPherson, Lynn Smith-Lovin, and James M. Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27:415–444, 2001.

Carl D. Meyer. *Matrix analysis and applied linear algebra.* SIAM, 2000.

Christopher J. Mruk. *Self-esteem research, theory, and practice: Toward a positive psychology of self-esteem.* Springer Publishing Company, 2006.

Ted O'Donoghue and Matthew Rabin. Choice and procrastination. *Quarterly Journal of Economics*, 116:121–160, 2001.

Ted O'Donoghue and Matthew Rabin. Procrastination on long-term projects. *Journal of Economic Behavior & Organization*, 66(2):161–175, 2008.

Emre Ozdenoren, Stephen W. Salant, and Dan Silverman. Willpower and the optimal control of visceral urges. *Journal of the European Economic Association*, 10(2): 342–368, 2012.

Keith T. Poole and R. Steven Daniels. Ideology, party, and voting in the us congress, 1959–1980. *American Political Science Review*, 79(2):373–399, 1985.

Milton E. Rosenbaum. The repulsion hypothesis: On the nondevelopment of relationships. *Journal of Personality and Social Psychology*, 51(6):1156, 1986.

Ariel Rubinstein. Economics and psychology? the case of hyperbolic discounting. *International Economic Review*, 44(4):1207–1216, 2003.

Thomas C. Schelling. Models of segregation. *The American Economic Review*, 59(2): 488–493, 1969.

Ross Sheldon et al. *A first course in probability.* Pearson Education India, 2002.

Ramadhar Singh and Soo Yan Ho. Attitudes and attraction: A new test of the attraction, repulsion and similarity-dissimilarity asymmetry hypotheses. *British Journal of Social Psychology*, 39(2):197–211, 2000.

Lones Smith and Peter Sørensen. Pathological outcomes of observational learning. *Econometrica*, 68(2):371–398, 2000.

Kenneth Sole, Janos Marton, and Harvey A. Hornstein. Opinion similarity and helping: Three field experiments investigating the bases of promotive tension. *Journal of Experimental Social Psychology*, 11(1):1–13, 1975.

John Stachurski. *Economic dynamics: theory and computation*. MIT Press, 2009.

Barry M. Staw. The escalation of commitment to a course of action. *Academy of management Review*, 6(4):577–587, 1981.

Daniel W. Stroock. *Probability theory: an analytic view*. Cambridge university press, 2010.

Irina Suanet and Fons J.R. Van de Vijver. Perceived cultural distance and acculturation among exchange students in russia. *Journal of Community & Applied Social Psychology*, 19(3):182–197, 2009.

Tomomi Tanaka, Colin F. Camerer, and Quang Nguyen. Risk and time preferences: linking experimental and household survey data from vietnam. *The American Economic Review*, 100(1):557–571, 2010.

Richard Thaler. Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, 1(1):39–60, 1980.

Zannie Giraud Voss, Daniel M. Cable, and Glenn B. Voss. Organizational identity and firm performance: What happens when leaders disagree about "who we are?". *Organization Science*, 17(6):741–755, 2006.