Games and Economic Behavior 142 (2023) 765–793

Contents lists available at ScienceDirect

Games and Economic Behavior

journal homepage: www.elsevier.com/locate/geb

On the transmission of guilt aversion and the evolution of trust ☆

Sebastiano Della Lena [a], Elena Manzoni [b,*], Fabrizio Panebianco [c]

[a] Department of Economics, Monash University, Australia
[b] Dipartimento di Scienze Economiche, Università degli Studi di Bergamo, Italy
[c] Dipartimento di Economia e Finanza, Università Cattolica del Sacro Cuore, Milano, Italy

ARTICLE INFO

JEL classification:
C72
D3
D80
Z10

Keywords:
Trust and trustworthiness
Guilt aversion
Cultural transmission
Psychological games

ABSTRACT

This paper studies the evolution of trust and trustworthiness by modeling the intergenerational transmission of guilt aversion. The results depend both on features of strategic interaction and on parental transmission. We show that if there is complete information of opponents' traits, independent of parenting style, the share of high-guilt agents in society weakly increases over time, and trust and trustworthiness are maximized. Moreover, when traits are not observable, different levels of guilt always coexist, and trust and trustworthiness might also increase when parents have imperfect empathy in the transmission of traits or if there is homophily in society.

1. Introduction

"(Religious) guilt could now appear not only as an occasional concomitant, but as an integral part of all culture, of all conduct in a civilized world, and finally, of all structured life in general. And thereby the ultimate values which this world offered have seemed burdened with the greatest guilt." [Max Weber 2013: 354]

Psychologists have long known that emotions influence human behavior and individual responses to positive and negative events. Recently, also the economic literature has recognized that emotions matter for strategic behavior (Battigalli and Dufwenberg, 2022). However, the formation and endogenous evolution of the personal traits that affect emotional responses has never been formally investigated.

☆ We thank Giuseppe Attanasi, Pierpaolo Battigalli, Alberto Bisin, Francesco Cinnirella, Martin Dufwenberg, Werner Güth, Weijia Li, Amrish Patel, Alexander Sebald, and participants to the 1st CoCoLab Workshop in Nice, the Vele Workshop in Verona, and the Australasian ASREC Conference in Melbourne. Sebastiano Della Lena and Elena Manzoni gratefully acknowledge fund from MIUR for the project PRIN 2017K8ANN4: *"New approaches to Political Economy: from methods to data"*. Sebastiano Della Lena gratefully acknowledges fund from FWO-foundation for the project *"Diffusion of Misinformation in Social Networks"* (id. 42933).

* Corresponding author.
E-mail addresses: sebastiano.dellalena@monash.edu (S. Della Lena), elena.manzoni@unibg.it (E. Manzoni), fabrizio.panebianco@unicatt.it (F. Panebianco).

https://doi.org/10.1016/j.geb.2023.09.012
Received 8 March 2023

Available online 5 October 2023

Trust and trustworthiness are of fundamental importance for different economic outcomes (Guiso et al., 2006) and, as pointed out by Alesina and Giuliano (2015), generalized trust towards others is "the most studied psychological trait."[1] Also, there is evidence that trust attitudes are transmitted across generations (Guiso et al., 2008a; Algan and Cahuc, 2010; Dohmen et al., 2012) and are positively correlated with religious sentiments (Tan and Vogel, 2008).

In this paper, we take a step back and investigate the emotional foundations of trust attitudes, focusing on *guilt aversion* modeled as a belief-dependent preference: individuals have a disutility—i.e., experience guilt—when disappointing others. As shown by both the theoretical and experimental literature, guilt aversion fosters trust and trustworthiness (Charness and Dufwenberg, 2006; Attanasi et al., 2016, 2019).[2]

In particular, we propose a model that analyzes how, depending on the features of the strategic interaction and the values of a society, the psychological trait of guilt aversion is transmitted from parents to children, shapes the dynamics of play in a Trust (Mini-) Game,[3] and consequently determines the evolution of trust and trustworthiness in the society. The paper predicts that when traits are observable, as in small communities, the share of high-guilt agents in society weakly increases over time, and trust and trustworthiness are maximized. Conversely, when traits are not observable, as in large anonymous societies, different levels of guilt coexist, and trust and trustworthiness may increase only if parents are imperfectly empathic in transmitting traits and if there is homophily in society.

The formal analysis of the effects of emotions on strategic behavior has been developed by what is called *psychological game theory* (Geanakoplos et al., 1989; Battigalli and Dufwenberg, 2009; Battigalli et al., 2019a, and references therein). Psychological game theory provides a way of showing how emotions play a role in planning strategies and choosing actions, by allowing individual utility to depend not only on outcomes but also on (one's own and others') beliefs. In detail, individual's emotional response to the strategic environment and to the co-players' actions is modeled with the introduction of psychological traits—e.g., guilt aversion (Battigalli and Dufwenberg, 2007; Battigalli et al., 2019a), frustration and anger (Battigalli et al., 2019b), or self-esteem (Mannahan, 2019). In the context of the Trust Game, Attanasi et al. (2019, 2022) show how, introducing guilt aversion modeled as a belief-dependent preference, the theoretical predictions fit better with experimental evidence than the standard models do.

The issue of where these psychological traits stem from has not yet been analyzed in psychological game theory. However, in the psychological literature, there is evidence that personality traits are formed in the early stages of human development (e.g., Bandura and Walters, 1963; Erikson, 1993, 1994). In particular, the psychological trait of guilt aversion is already present in children aged 10–12 years old (Ferguson et al., 1991), and its development occurs during childhood (Kochanska et al., 2002).

In this respect, our contribution is to endogenize and study the dynamics of guilt aversion, taking into account: (*i*) different styles of parental transmission of traits; (*ii*) the effects of population shares on parental socialization efforts—i.e., cultural substitution vs. complementarity; (*iii*) features of the strategic environment, such as the observability (or not) of individual traits and the presence (or not) of assortativity in the matching.

Specifically, we consider a population partitioned into two groups characterized by different levels of guilt aversion (*H*igh or *L*ow). Two players, *A* and *B*, are matched, and player *A* (she) is endowed with 2 monetary units. She has to decide whether to dissolve the partnership, choosing *O*ut and dividing the money equally with player *B*, or to remain in the partnership by choosing *I*n. If she remains *I*n, the money is doubled and transferred to player *B* (he), who can, in turn, decide to *S*hare it equally with *A* or to *T*ake the whole amount. We define *trust* as the share of agents choosing *I*n and *trustworthiness* as the share of agents choosing *S*hare.

Guilt aversion may play an important role in players' choices. In this paper, consistent with insights from the evolutionary psychology of emotions (Haselton and Ketelaar, 2006), we focus on the case in which players may experience guilt only when in role *B*.[4] Following Battigalli et al. (2019a), *B* feels guilty if he believes he "lets player *A* down." Given the nature of the game, player *A* can only be disappointed by player *B* after *I*n, as this is the only case in which player *B* has an active choice. Moreover, *A* may be disappointed only when *B* decides to *T*ake. In this case, *B*'s guilt is proportional to *B*'s guilt sensitivity and (beliefs about) *A*'s expectation of receiving back half of the monetary units.

We assume that guilt aversion is a trait that is acquired during an early socialization process and that it does not change during life, once acquired. We use the standard model of the intergenerational transmission of cultural traits (Bisin and Verdier, 2001, 2011).

In detail, each agent acquires their trait during childhood through two main mechanisms: (*i*) *vertical* (direct) *transmission*, in which each parent exerts a *socialization* effort to transmit their trait to the child; and (*ii*) *oblique transmission*, in which each child may acquire a random trait from the adult population in the society. This induces the dynamics of shares of agents displaying each of the two levels of guilt aversion. These dynamics crucially depend on how parents choose their socialization efforts.

[1] Many empirical studies show that culture and, in particular, social capital—which is mostly represented by trust and trustworthiness—has a huge impact on different economic phenomena, such as economic growth (Knack and Keefer, 1997), size of firms (La Porta et al., 1997), financial development (Guiso et al., 2004, 2008b), the quality of institutions (Tabellini, 2008a, 2010).

[2] Note also that guilt aversion is the predominant psychological trait found in the data, particularly in Trust Games (Attanasi et al., 2019, 2022; Bellemare et al., 2017, 2019).

[3] The Trust Minigame (named after Attanasi et al., 2016), is a binary-choice version of the Trust Game, similar to the one proposed by Kreps (1990). Henceforth, we will always refer to it simply as Trust Game.

[4] We acknowledge that agents might also feel guilty when in role *A*. Thus, in Section 7 we will briefly discuss the consequences of having players who experience guilt also when they are in role *A*.

The parental effort choice is determined by the relative advantage parents think their children will have if they are part of their own group, as opposed to the opposite group—called *cultural intolerance* in the literature (see Bisin and Verdier, 2001). We model cultural intolerance by considering parents who care about different aspects of children's future utility when choosing the socialization effort. In detail, we define as *materialistic* the parents who exert efforts that only depend on how different levels of guilt sensitivity in the two groups induce higher or lower material payoffs. Then, we also consider societies in which parents care about the fact that children may experience guilt on top of material payoffs, and, thus, socialization efforts are chosen by comparing the expected psychological payoffs gained by the children with high or low guilt sensitivity during the interactions. In particular, we consider both the cases in which all the parents are *perfectly empathic*—in the sense that they compute psychological payoffs of each group using the correct psychological parameter—, and the case in which they are all *imperfectly empathic*[5]—namely they just use their own psychological parameter to compute the psychological payoffs of both groups. Note that imperfect empathy may arise, for example, if parents are moved by the fear of some moralizing god (Norenzayan, 2013). Indeed, thinking about large monotheistic religions, when parents believe that a supernatural entity may punish them or their children for misconduct, or lack of cooperation, they will evaluate children's utility according to their belief—i.e., with imperfect empathy. Since in the above three cases (materialism, perfect empathy, and imperfect empathy) parents differ in what they consider relevant to socializing children, we name them as different *parental values*.

It is a well-known result (Bisin and Verdier, 2001) that, whenever cultural intolerance is independent of population share, the socialization effort displays *cultural substitution*—i.e., it decreases in own group's population share—and, this phenomenon generally leads to long-run cultural heterogeneity (Bisin and Verdier, 2011). In our framework, as previously discussed, cultural intolerance depends on the expected utility of the adult age game (Trust Game) evaluated differently depending on the parental values and, thus, it may be affected by the shares of high and low guilt agents in the society. We show the way in which cultural intolerance reacts to population shares may mitigate the standard baseline substitutability that arises from the will to transmit their own trait. In particular, the socialization effort displays *cultural complementarity*—i.e., it is increasing in the own group's population share—if and only if the cultural intolerance reacts positively to changes in the size of the own group and it is elastic, so that, it is reactive enough to population shares to overcome the baseline substitutability.

We start by considering societies where the individual traits are *observed* by all the agents and, thus, every individual observes their partner's trait. Then, only two equilibrium outcomes are possible: either the full cooperation path is observed; or player A chooses Out at the beginning of the game. In both cases, agents do not experience any guilt and, as a consequence, the social dynamics is independent of the parental values. Then, given complete information about the matching, individuals from the group with high guilt are weakly more often in pairs in which the cooperative path is chosen and, therefore, their average payoff is weakly higher than the one of low-guilt individuals. Thus, in the long run, the share of high-guilt individuals weakly increases as the levels of trust and trustworthiness do. We show that, in this case, there is always cultural substitution which is still compatible with long-run cultural homogeneity.

However, in large societies, individual traits are mostly *not observable*. In this case, the equilibrium path may have agents A trusting their partners and some of their matched partners B betraying this trust. The existence of this equilibrium path has two effects. First, guilt may be experienced in equilibrium by agents from the low guilt group, hence, parental values matter. Second, low-guilt individuals have high material payoffs when they betray their partners' trust. We find that, if parents are materialistic or perfectly emphatic, the level of trustworthiness in the society weakly decreases over time, whereas, if parents are imperfectly empathic, it may decrease or increase depending on the maximum level of guilt present in the society. Moreover, we find that cultural complementarity may arise in equilibria in which betrayed trust is observed with positive probability, depending on parental values and guilt sensitivities.

In the paper, we also show that social structure impacts the dynamics of trust and trustworthiness. We acknowledge that the matching among agents may be assortative —i.e., the interaction pattern displays *homophily*. Indeed, assortativity in forming social contacts has commanded a lot of attention and has been largely observed in many social contexts (e.g., Coleman, 1958; Currarini et al., 2009). In such a case, to separate the effects, we focus only on societies with materialistic parents, showing that, if homophily is low, the levels of trust and trustworthiness in the society weakly decrease over time, whereas if homophily is high, the levels of trust and trustworthiness increase (weakly decrease) for high (low) population shares.[6]

Related literature This paper contributes, on top of the psychological game theory literature discussed above, to the theoretical literature of cultural transmission that analyzes the interactions between the intergenerational transmission of traits and the strategic environment. In detail, Bisin et al. (2004) and Tabellini (2008b), study the evolution of cooperation in a Prisoner's Dilemma, focusing on complete vs. incomplete information about the matching and the spatial interaction among agents, respectively. Della Lena and Dindo (2023) study the different dynamics of acculturation when agents interact in strategic environments with either complements or substitutes.

[5] Note that, we refer to *imperfectly empathic* parents in a slightly different way than in Bisin and Verdier (2001). Indeed, we refer to the incapacity of parents to use the correct guilt parameter to valuate their psychological utility, which can be possibly generated by their inability to fully empathize with agents who hold different preferences. Their standard assumption of imperfect empathy (i.e., paternalistic altruism), that always leads parents to attempt to socialize their children to their own trait, always holds in the paper.

[6] We limited our analysis to the materialistic parents case as the standard analysis without homophily already show that imperfect empathy has a positive effect on trust and trustworthiness. Therefore, we prefer to investigate the effect of homophily on trust and trustworthiness without the confounding interaction with imperfect empathy.

Lastly, in Guiso et al. (2008a) and Okada (2020) agents interact in a Trust Game, as in this paper. In particular, in Guiso et al. (2008a) parents transmit their beliefs about the trustworthiness of others, whereas in Okada (2020) parents transmit the psychological benefit to have a "good conduct". The main difference is that, in our work, parents transmit their level of guilt sensitivity which induces trustworthiness and consequently trust and the beliefs are determined in equilibrium.

The paper also relates to the literature about the indirect evolution of preferences (Güth and Yaari, 1992; Dekel et al., 2007) and its applications in the context of the Trust Game (e.g., Güth and Kliemt, 1994). Our approach is different on two dimensions: (*i*) to the best of our knowledge, we are the first to study the evolution of a psychological trait that affects belief-dependent preferences, instead of an exogenous psychological preference[7]; and (*ii*) the evolution of traits is not due to a mechanical biological process but to a purposeful parental transmission. The differences are not only methodological but also substantive. If we analyze trust taking into account guilt aversion, the theoretical predictions better match experimental evidence, so that the model provides a more appropriate description of the co-evolution of personal traits and behavior. Moreover, in our model, long-run results crucially depend on the characteristics of parental preferences in the purposeful transmission of traits. This mechanism cannot be fully captured by a (reduced-form) indirect evolution model.

The paper also informs the literature about economics of religions (see Iannaccone, 1998; Iyer, 2016; Carvalho et al., 2019, and the references therein). Indeed, as stressed by Weber (2013), the sentiment of guilt is a crucial feature of religion and an integral part of many cultures. Also, Guiso et al. (2003, 2006) show that religious people are generally more trusting. Since religiosity may increase guilt attitudes, studying the evolution of trust—stemming from its emotional foundation in guilt aversion and the consequent intergenerational cultural transmission—allows us to better understand the conditions that favor the diffusion of religiosity and its relation with prosocial behavior in a society (e.g., Shariff and Norenzayan, 2007; Norenzayan and Shariff, 2008; Levy and Razin, 2012, 2014; Bentzen, 2019). Papers like Bisin and Verdier (2000), Carvalho (2013) Patacchini and Zenou (2016), and Bisin et al. (2023), explicitly model the transmission of religion in a cultural transmission fashion. We are the first to consider the intergenerational transmission and dynamics of a guilt sensitivity trait which can be interpreted as an religiosity-induced trait and to take into account its effect on agents' decisions and on the transmission process.

The paper is structured as follows. Section 2 presents the notation and the main features of the model: the Trust Game; the Social Dynamics; and the Parental Values. In Section 3, we show how parental socialization effort reacts to changes in population shares. Sections 4 and 5 characterize the equilibrium strategies of the adult age (trust) game and the implied social dynamics for different parental values with complete and incomplete information, respectively. Section 6 analyzes the presence of homophily in interactions. Section 7 discusses alternative assumptions and the implications of the model for the dynamics of religiosity. Section 8 concludes the paper. All the proofs of the propositions in the main text are in Appendix A. Appendix B characterizes the equilibria of the Trust Game.

2. The model

We consider a society that, at each time $t \in \mathbb{N}_0$, is composed by a cohort of agents of mass 1 who are alive only at time t. Each agent has belief-dependent preferences which display guilt aversion, and, at time t, she/he is randomly matched infinitely many times with other agents to play what we called adult age game, which in our model is a Trust Game described in Section 2.1, maximizing her/his *instantaneous* expected utility. Each agent, before dying, asexually gives birth to one child (alive at $t+1$). Each parent, given the outcome of their strategic interactions, chooses a socialization effort to transmit their own level of guilt aversion to the offspring (Section 2.2). Then parents die and the process starts again with a new generation.

2.1. Trust game

We assume that the adult age game played in the population is a Trust Game (Berg et al., 1995; Attanasi et al., 2016). In the Trust Game player A (she)[8] receives an amount 2 and has to decide whether to split this amount evenly with player B (he) by going Out (O) or to transfer the whole amount to B by choosing In (I). If player A transfers it, the amount is doubled, and player B can decide whether to Share (S) it evenly with player A or to Take (T) it all for himself. Fig. 1 shows the Trust Game with material payoffs.

We assume that both players are randomly drawn from a population with heterogeneous levels of guilt aversion, described below. We assume that guilt is role-dependent, namely an individual may experience guilt only if he is drawn to play in the role of player B. This is a simplifying assumption, which is however consistent with insights from the evolutionary psychology of emotions, where it is argued that when a single emotion operates in different situations its consequences are affected by contextual cues (Haselton and Ketelaar, 2006). Therefore, player A's utility is not affected by guilt. Let us denote A's strategies with $s_A \in \{O, I\}$ and B's strategies with $s_B \in \{T, S\}$.[9] We assume that player A's utility is $u_A(s) = m_A(s)$, where $m_A(s)$ denotes material payoff of agents in role A, after the terminal history induced by the strategy profile $s := (s_A, s_B)$.

We model guilt aversion as Battigalli and Dufwenberg (2007) and Battigalli et al. (2019a). In order to model the effects of player B's guilt aversion on his behavior we need to define players' first- and second-order beliefs. As in Battigalli et al. (2019a), we assume

[7] See the discussion about Psychological Game Theory above.

[8] For simplicity of exposition we think of player A being female and player B being male. Of course, this is just a convention, as every male and female individual in each population will play both roles with the same probability.

[9] Note that we do so with a slight abuse of notation, as we call T (S) not only B's action Take (Share) but also B's strategy Take (Share) if In.

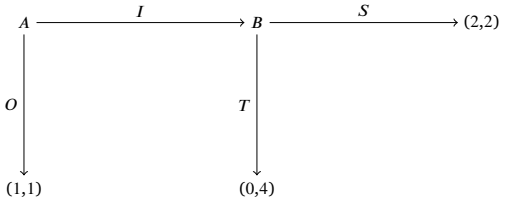


Fig. 1. Trust Game with material payoffs.

that players' plans are logically distinct from their behavior, and as a consequence we can meaningfully define players' beliefs on own strategies. In particular, let $\boldsymbol{\alpha}_A, \boldsymbol{\alpha}_B$ denote players' first-order beliefs and $\boldsymbol{\beta}_A, \boldsymbol{\beta}_B$ denote players' second-order beliefs.[10] First-order beliefs are players' (probabilistic) beliefs on primitive uncertainty, such as own and partner's strategies; second-order beliefs are subjective probability measures about primitive uncertainty and about the partner's beliefs. In the analysis of the game we will introduce some features of these beliefs, for which we introduce a specific notation.

Player B, if guilt averse, suffers from guilt when he believes he is letting player A down. As a matter of notation let define, for a generic x, $[x]^+ := \max\{0, x\}$. With this, A's **disappointment** is defined as

$$D_A(\boldsymbol{s}, \boldsymbol{\alpha}_A) := \left[\mathbb{E}_{\boldsymbol{\alpha}_A}[\tilde{m}_A] - m_A(\boldsymbol{s})\right]^+, \tag{1}$$

which is the difference between A's expected value of her material payoff and her realized material payoff given the implemented strategy profile $\boldsymbol{s}$, when the difference is positive. Note that $\tilde{m}_A$ is, from A's point of view, a random variable, as it depends on B's choices as well. As in Battigalli and Dufwenberg (2007), we assume that player B feels guilty only for that component of player A's disappointment that is due to his behavior. Player B has belief-dependent preferences over material payoffs represented by the following psychological utility function:

$$u_B(\boldsymbol{s}, \boldsymbol{\alpha}_A; \theta) = m_B(\boldsymbol{s}) - \theta \left[D_A(\boldsymbol{s}, \boldsymbol{\alpha}_A) - \min_{s'_B} D_A(s_A, s'_B, \boldsymbol{\alpha}_A) \right], \tag{2}$$

where: $\theta \in \mathbb{R}_+$ is player B's **guilt sensitivity** — i.e., his level of guilt aversion—; $\min_{s'_B} D_A(s_A, s'_B, \boldsymbol{\alpha}_A)$ is the minimum level of disappointment B can deliver to A; and $\left[D_A(\boldsymbol{s}, \boldsymbol{\alpha}_A) - \min_{s'_B} D_A(s_A, s'_B, \boldsymbol{\alpha}_A)\right]$ is the component of player A's disappointment for which player B is responsible and feels guilty. Note that, in a two-period game without chance move, this model is equivalent in terms of best replies to a model in which player B experiences a guilt proportional to the full disappointment of player A, and not to his own contribution to it. However, the two models differ in terms of B's utility after Out, which is relevant in our model, as it affects the average lifetime utility of players.

Let $\alpha_A^S = \mathbb{P}_A[S]$ be the unconditional probability that player A assigns to player B Sharing, and $\alpha_A^I = \mathbb{P}_A[I]$ the unconditional probability that player A assigns to herself going In. These are features of A's first-order belief. Given that we assume that players' plans are logically distinct from their behavior (Battigalli et al., 2019a), players need not to be consistent with them. Therefore α_A^I, which is player A's belief on his In choice, can take values different from 0 or 1, even though in the analysis we only focus on pure strategy equilibria. Hence, player A can be disappointed not only after terminal history (I, T), but also after terminal history O (in case she planned to go In to obtain a higher payoff, but failed to do so). However, the only case in which player B can experience guilt is when player A's disappointment is caused by player B's choice. Therefore, in this model, player B can feel responsible for A's disappointment only after the terminal history (I, T). In this case $m_A(I, T) = 0$, and player A's disappointment is

$$D_A(\boldsymbol{s}, \boldsymbol{\alpha}_A) = \left[\mathbb{E}_{\boldsymbol{\alpha}_A}[\tilde{m}_A] - 0\right]^+ = (1 - \alpha_A^I) \cdot 1 + \alpha_A^I \cdot 2\alpha_A^S.$$

Moreover, given that player B could grant player A her maximum payoff by choosing to Share, $\min_{s'_B} D_A(s_A, s'_B, \boldsymbol{\alpha}_A) = 0$. Thus, the psychological utility of (I, T) for player B (expressed as a function of player A's first-order belief) is

$$u_B(I, T, \boldsymbol{\alpha}_A; \theta) = 4 - \theta\left(1 - \alpha_A^I + 2\alpha_A^S \alpha_A^I\right).$$

Fig. 2 represents the Trust Game with our specification of psychological utilities.

Note that player B does not know A's first-order belief $\boldsymbol{\alpha}_A$, and therefore he chooses his strategy based on his second-order belief $\boldsymbol{\beta}_B$.[11]

[10] Later in the paper we will introduce specific notation for the features of first- and second-order beliefs that are relevant for our analysis.

[11] The features of $\boldsymbol{\beta}_B$ that are relevant for the analysis are the expectations of α_A^S and α_A^I that player B has, conditional on player A having chosen In. We let $\beta_B(S) = \mathbb{E}[\alpha_A^S | In]$, and $\beta_B(I) = \mathbb{E}[\alpha_A^I | In]$. Note that the latter may be smaller than 1, because B may think that player A goes In assigning positive probability to a mistake.

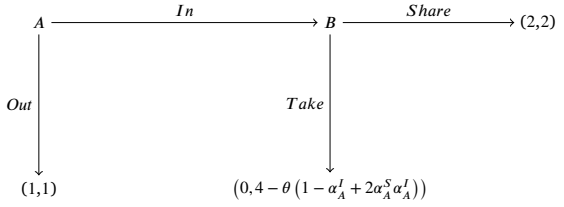


Fig. 2. Trust Game with psychological utilities.

2.2. Social dynamics

We now introduce the cultural composition of the society and discuss the social dynamics implied by the outcome of strategic interactions in the adult age game.

Each agent belongs to one of two homogeneous cultural groups, where $C := \{L, H\}$ is the set of groups. Each agent has guilt sensitivity in $\Theta := \{\theta^L, \theta^H\} \in \mathbb{R}^2_+$, where Θ is common knowledge. Without loss of generality, we assume $\theta^L \leq \theta^H$, so that L is the low guilt group and H the high guilt one. At any given $t \in \mathbb{N}_0$, let q_t^i be the measure of group $i \in C$ in the society.

As discussed above, agents play infinitely many times the Trust Game described in Section 2.1. Each time the game is played, agents are matched with a partner randomly drawn from the whole population and they play in both roles (A, B) with the same probability.

At the end of the strategic interactions, agents observe the frequency of actions played in the society by agents of each group and, thus, assuming common knowledge of the game form, they can compute the average payoffs.

With a little abuse of notation, we let $i \in C$ denote both the group and the representative agent belonging to that group. Each parent $i \in C$ chooses a socialization effort $\tau_t^i \in [0,1]$ to transmit own trait to their own child.[12] In order to choose an effort each parent must have an expectation of the utility they derive from having a child with their own guilt sensitivity as opposed to a child with a different one. At each time t, each parent $i \in C$ has a conjecture about the future population shares at $t+1$. We assume *adaptive expectations* along the paper, so that parents form conjectures about population shares at $t+1$ using the observed population shares at t. This implies that all agents share the same conjecture. Formally, for each $i \in C$, $\mathbb{E}_{\delta_{q_t^i}}[q_{t+1}^i] = q_t^i$, where the common conjecture $\delta_{q_t^i}$ is the Dirac measure at q_t^i.[13] This means that parents do not internalize changes in the population shares due to the cultural dynamics we describe below. For each $i \in C$ and $j \neq i$, let V_t^{ii} and V_t^{ij} be the utility each parent i at time t expects from having a child in group i or j, respectively. In Section 2.3, we discuss how these utilities are formed, allowing both for the case in which parents just care about own child's material payoffs, and for the case in which they also care about psychological utilities.

We now describe the transition probabilities which characterize the cultural transmission process (Bisin and Verdier, 2001). In detail, each parent directly socializes the child to own trait with a probability equal to their own socialization effort τ_t^i – *vertical socialization*. If this socialization fails, the child randomly takes a trait from the population – *oblique socialization*. Therefore, the probability that, at time t, a child of parent i acquires trait i is given by

$$p_t^{ii} := p(\tau_t^i; q_t^i) = \tau_t^i + (1-\tau_t^i)q_t^i, \tag{3}$$

Analogously, $p_t^{ij} = (1-\tau_t^i)(1-q_t^i)$.

Each parent $i \in C$, given q_t^i, chooses the effort τ_t^i that maximizes their subjective expected utility

$$\mathbb{E}_{p_t^{ii}}(u_t^i) = p_t^{ii}V_t^{ii} + (1-p_t^{ii})V_t^{ij} - \frac{1}{2}(\tau_t^i)^2, \tag{4}$$

where $\frac{1}{2}(\tau_t^i)^2$ is the cost of socialization. Indeed, with probability p_t^{ii} they get a child with their own trait and gain V_t^{ii}, while with probability p_t^{ij} they get a child with a different trait and gain V_t^{ij}. Let $\Delta V_t^i := V_t^{ii} - V_t^{ij}$, and similarly $\Delta V_t^j := V_t^{jj} - V_t^{ji}$. These are referred to in the literature as *cultural intolerance* of group i and j, respectively. Solving the parental optimization problem we get, for each $i \in C$,

$$\tau_t^i := \tau(q_t^i, \Delta V_t^i) = \left[(1-q_t^i)\Delta V_t^i\right]^+. \tag{5}$$

Note first that, if parents of group i expect to receive a higher utility from having a child with a different trait than a child with their own trait, they do not exert any effort. The higher the cultural intolerance ΔV_t^i the higher the effort, given that parents think that being of their own trait is the more profitable for their children the higher ΔV_t^i. Moreover, it is trivial to see that, when ΔV_t^i is independent of q_t^i, parents of group i increase their socialization effort when their population share decreases. This property is

[12] We refer to agents as parents when we discuss their socialization efforts, i.e., when they take actions related to their role as parents. Moreover, we refer to both parents and children with the singular *they*.

[13] Note that this assumption implies that at steady states conjectures about population shares are always correct. We refer to Della Lena and Panebianco (2021) to understand the effect of wrong conjectures about population shares in a cultural transmission framework.

known in literature as *cultural substitution*. In Section 3 we study what happens, instead, when cultural intolerance is endogenously determined by population shares, and this endogeneity is of paramount importance for our analysis.

The population dynamics is given by

$$q^i_{t+1} = q^i_t \cdot p^{ii}_t + (1 - q^i_t) \cdot p^{ji}_t.$$

Using the continuous-time approximation we get

$$\dot{q}^i_t = q^i_t(1 - q^i_t)(\tau^i_t - \tau^j_t), \tag{6}$$

which, when both efforts are positive, reads

$$\dot{q}^i_t = q^i_t(1 - q^i_t)\Big((1 - q^i_t)\Delta V^i_t - q^i_t \Delta V^j_t\Big). \tag{7}$$

Lastly, when one effort is zero, the dynamics is trivial.

2.3. Parental values

Parents may care about different aspects of their children's life while forming subjective expected utility from having children with different traits. Depending on the cultural features of the society parents live in, some of these aspects may be prevalent.[14] For our purposes, we consider both parents who care only about material payoffs and parents who care also about children's psychological utilities.

Let $\boldsymbol{s}_t := (s^i_t)_{i \in C}$ be the equilibrium strategy profile of the adult age game at time t.[15] For each $i \in C$ and $t \in \mathbb{N}_0$, define $\bar{m}^i_t$ as the average material payoff of agent $i \in C$ in the stage game. Given our assumptions about the matching, this is equal to the average lifetime material payoff agent $i \in C$ experiences. Namely,

$$\bar{m}^i_t := \bar{m}^i(\boldsymbol{s}_t) = \frac{1}{2} m^i_A(\boldsymbol{s}_t) + \frac{1}{2} m^i_B(\boldsymbol{s}_t),$$

where, $m^i_A(\boldsymbol{s}_t)$ is the lifetime payoff gained by agents when playing in role A, and $m^i_B(\boldsymbol{s}_t)$ is the one gained when playing in role B. Similarly, we define the average psychological utility of agent $i \in C$ in the adult age game as

$$\bar{u}^i_t := \bar{u}^i(\boldsymbol{s}_t; \theta^i) = \frac{1}{2} u^i_A(\boldsymbol{s}_t; \theta^i) + \frac{1}{2} u^i_B(\boldsymbol{s}_t; \theta^i),$$

where $u^i_A(\boldsymbol{s}_t; \theta^i)$ is the lifetime utility experienced by agents when playing in role A, whereas $u^i_B(\boldsymbol{s}_t; \theta^i)$ is the one experienced playing in role B. We distinguish between materialistic parents and those who care also about children's psychological utilities. The latter can be further classified as perfectly and imperfectly empathic parents. Specifically,

(M) **Materialistic parents**. These parents care only about their own children's material payoffs, so that $V^{ii}_t = \bar{m}^i_t$ and $V^{ij}_t = \bar{m}^j_t$. Given that $\mathbb{E}_{\delta_{q^i_t}}[q^i_{t+1}] = q^i_t$, parents expect the average material payoff of each group at $t+1$ to be the same as the average material payoff of that group at time t, that is, for all $i, j \in C$, $\mathbb{E}_{\delta_{q^i_t}}[\bar{m}^i_{t+1}] = \bar{m}^i_t$, $\mathbb{E}_{\delta_{q^i_t}}[\bar{m}^j_{t+1}] = \bar{m}^j_t$. Thus, $\Delta V^i_t = \frac{\bar{m}^i_t - \bar{m}^j_t}{\bar{m}^i_t + \bar{m}^j_t} = -\Delta V^j_t$, which implies that only one of the two efforts is positive and, thus, the social dynamics, for each is given by

$$\begin{cases} \dot{q}^i_t = q^i_t(1 - q^i_t)^2 \left(\bar{m}^i_t - \bar{m}^j_t\right) & if \quad \bar{m}^i_t > \bar{m}^j_t \\ \dot{q}^i_t = (q^i_t)^2(1 - q^i_t) \left(\bar{m}^i_t - \bar{m}^j_t\right) & if \quad \bar{m}^i_t < \bar{m}^j_t \end{cases}, \tag{8}$$

i.e., $\dot{q}^i_t$ is positive when $\bar{m}^i_t > \bar{m}^j_t$ and negative when the opposite relation holds.

(PE) **Perfectly empathic parents.** These parents care about their own children's psychological utility. When parents evaluate the expected utility of their own children being of a given cultural group, they are able to perfectly identify themselves with agents of that group. Therefore, they use the correct guilt sensitivity to compute the psychological utility experienced by agents of that group. Given that $\mathbb{E}_{\delta_{q^i_t}}[q^i_{t+1}] = q^i_t$, it follows that, for all $i, j \in C$, $\mathbb{E}^j_{\delta_{q^i_t}}[\bar{u}^i_{t+1}] = \mathbb{E}^i_{\delta_{q_t}}[\bar{u}^i_{t+1}] = \bar{u}^i_t$. Thus we have that $\Delta V^i_t = \frac{\bar{u}^i_t - \bar{u}^j_t}{\bar{u}^i_t + \bar{u}^j_t} = -\Delta V^j_t$, which implies that only one of the two efforts is positive and, thus, the dynamics is given by

$$\begin{cases} \dot{q}^i_t = q^i_t(1 - q^i_t)^2 \left(\bar{u}^i_t - \bar{u}^j_t\right) & if \quad \bar{u}^i_t > \bar{u}^j_t \\ \dot{q}^i_t = (q^i_t)^2(1 - q^i_t) \left(\bar{u}^i_t - \bar{u}^j_t\right) & if \quad \bar{u}^i_t < \bar{u}^j_t \end{cases}, \tag{9}$$

[14] These cultural features can be shaped for example by cultural leaders (see, for example, Verdier and Zenou, 2015; Prummer and Siedlarek, 2017; Verdier and Zenou, 2018).

[15] We focus here on the case in which the adult age game has a unique equilibrium strategy profile. In the characterization of the equilibria we will also provide selection criteria such that a unique equilibrium is selected at any parameterization.

i.e., $\dot{q}_t^i$ is positive when $\bar{u}_t^i > \bar{u}_t^j$ and negative when the opposite relation holds.

(*IE*) **Imperfectly empathic parents.** These parents care about their own children's psychological utilities but, when they evaluate the expected utility of a child of any cultural group, they use their own guilt sensitivity to compute the offspring's psychological utility. In detail, $\mathbb{E}^i_{\delta_{q_t^i}}[\bar{u}^i_{t+1}] = \frac{1}{2}u^i_A(\boldsymbol{s}_t;\theta^i) + \frac{1}{2}u^i_B(\boldsymbol{s}_t;\theta^i) = \bar{u}^i_t$ as before. When $j \neq i$, instead,

$$\hat{u}_t^j := \mathbb{E}^i_{\delta_{q_t^i}}[\bar{u}^j_{t+1}] = \frac{1}{2}u^j_A(\boldsymbol{s}_t;\theta^i) + \frac{1}{2}u^j_B(\boldsymbol{s}_t;\theta^i).$$

Then $\Delta V_t^i = \frac{\bar{u}_t^i - \hat{u}_t^j}{\bar{u}_t^i + \hat{u}_t^j} \neq -\Delta V_t^j$. As a consequence, both parental efforts may be positive, in which case the dynamics is given by (7), and it becomes

$$\dot{q}_t^i = q_t^i(1-q_t^i)\Big((1-q_t^i)(\bar{u}_t^i - \hat{u}_t^j) - q_t^i(\bar{u}_t^j - \hat{u}_t^i)\Big). \tag{10}$$

3. Cultural substitution and complementarity

In this section we analyze how parental socialization effort reacts to changes in population shares. Formally, we say that a socialization effort τ_t^i displays *cultural substitution* if $\frac{\partial \tau_t^i}{\partial q_t^i} < 0$, whereas it displays *cultural complementarity* if $\frac{\partial \tau_t^i}{\partial q_t^i} > 0$. We also define the elasticity of the cultural intolerance of group i with respect to the share of group j in population as $\varepsilon_{ij} := \frac{\partial \Delta V_t^i}{\partial q_t^j}\frac{q_t^j}{\Delta V_t^i}$. Recall that ΔV_t^i represents parent i's expectation on the relative material or psychological advantage of having a child belonging to own group. Then $\frac{\partial \Delta V_t^i}{\partial q_t^j}$ is a measure of the effect of a change of the population share of group j on the effort of parents i, passing through cultural intolerance. Therefore, the elasticity provides a measure of the responsiveness of the cultural intolerance to a change in the size of the opposite group.

Proposition 1. *Consider the optimal socialization effort in equation* (5)*. Then, for each* $i \in C$ *and* $j \neq i$*,*t

- *if* $\frac{\partial \Delta V_t^i}{\partial q_t^j} \geq 0$, τ_t^i *displays cultural substitution;*
- *if* $\frac{\partial \Delta V_t^i}{\partial q_t^j} < 0$, τ_t^i *displays cultural substitution if and only if* $\varepsilon_{ij} > -1$.

Recall from equation (5) that the optimal socialization effort is $\tau_t^i = \left[(1-q_t^i)\Delta V_t^i\right]^+$. Then, we see that, besides the effect of q_t on the cultural intolerance ΔV_t^i, there is always a baseline level of cultural substitution.

Proposition 1 shows that, when the cultural intolerance increases in the share of agents belonging to the other group—so that $\varepsilon_{ij} > 0$—, an additional motive of substitution, stemming from the change in the payoffs of the adult age game, comes into play and, thus, the socialization effort necessarily displays cultural substitution.

If, instead, the cultural intolerance decreases in the number of agents belonging to the other group—so that, $\varepsilon_{ij} < 0$—a complementarity between own group size and socialization effort arises. This happens whenever the average payoff of agents in group i is higher when interacting with agents of group j than with agents of the same group. The magnitude of this overall effect depends on how responsive cultural intolerance is to population shares. Thus, if the cultural intolerance of i is (negatively) elastic with respect to j's group size (i.e., $|\varepsilon_{ij}| > 1$), the complementarity effect dominates and, thus, socialization efforts display cultural complementarity. Conversely if the cultural intolerance is rigid with respect to the other group's size (i.e., $|\varepsilon_{ij}| < 1$), cultural substitution is displayed.

In the next sections, together with the analysis of the equilibria and of the population dynamics, we also describe how cultural intolerance depends on the specific parental value and on the level of guilt sensitivity, so as to analyze more in detail the cultural substitution/complementarity properties (see Corollary 1 and 2).

4. Complete information about the matching

We first study, as a benchmark, the case with complete information about the matching, in which traits are observable—i.e., both matched agents know which group their co-player belongs to and this is common knowledge. Complete information about the matching means complete information over the partner's guilt sensitivity θ.[16]

Since there are only two cultural groups in the society, from this section on we will refer to the population share of low guilt agents, q_t^L, as q_t and to the one of high guilt agents, $q_t^H = 1 - q_t^L$, as $1 - q_t$.

[16] We assume that player B's utility depends on player A's expectation given the matching. Player A could also form her expectation on her per-period payoff before the match is known. More than that, player B could feel guilt for letting player A down from her *initial* expectations, instead than letting her down from her expectations given the match. The analysis of the case in which there is complete information on the matching, but player B cares about player A's ex-ante disappointment is available from the authors upon request.

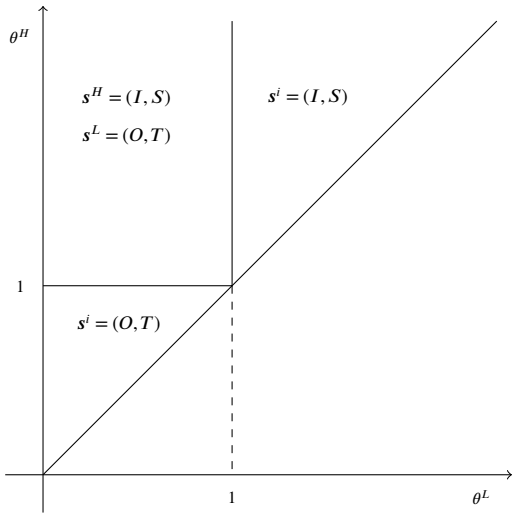


Fig. 3. Equilibrium behavior for agents of the two groups in the case of common knowledge of the matching.

We analyze the Bayesian Sequential Equilibria (BSE) in pure strategies of this Trust Game with role-dependent guilt aversion.[17]

Note that we have four possible types of match, and, in principle, four possible equilibria to be defined, indexed by the specific match ji—player A of group j, player B of group i—, with strategies $\boldsymbol{s}^{ji} := \{s_A^{ji}, s_B^{ji}\}$, and second-order beliefs $\boldsymbol{\beta}^{ji} := \{\beta_A^{ji}, \beta_B^{ji}\}$. With role dependent guilt aversion, only group i of player B matters, as it determines his guilt sensitivity θ^i. Hence, equilibrium strategies are just indexed by the group of player B, and for every $j, i \in C$, $\boldsymbol{s}^i := \boldsymbol{s}^{ii} = \boldsymbol{s}^{ji}$ {and $\boldsymbol{\beta}^i := \boldsymbol{\beta}^{ji} = \boldsymbol{\beta}^{ii}$}. Therefore, a BSE (in pure strategies), given a match in which the B player belongs to group i, is given by a (pure) strategy profile $\boldsymbol{s}^i = (s_A^i, s_B^i)$, and a corresponding profile of second-order beliefs $\boldsymbol{\beta}^i = (\beta_A^i, \beta_B^i)$, where second-order beliefs $\boldsymbol{\beta}^i$ are correct.[18,19]

We refer to Battigalli et al. (2019a) for the full characterization of the equilibria, and we simply report the (pure) strategy profiles that are compatible with equilibrium. These equilibrium strategy profiles, depending on the guilt parameter of the player B's θ^i, are:

- $\boldsymbol{s}^i \in \{(O, T)\}$, if $\theta^i < 1$;
- $\boldsymbol{s}^i \in \{(O, T), (I, S)\}$, if $\theta^i \in [1, 2]$;
- $\boldsymbol{s}^i \in \{(I, S)\}$, if $\theta^i > 2$.

This equilibrium characterization tells that if guilt aversion is low, the equilibrium path of play is the same as in the Nash equilibrium of the game without belief dependent preferences, and agents always choose to stay Out. On the contrary, as guilt aversion increases, also (I, S) becomes an equilibrium strategy profile. If guilt aversion is very high, then only (I, S) can be sustained in a BSE.

Note that, as just mentioned, for values of guilt aversion $\theta^i \in [1, 2]$ there is a multiplicity of strategy profiles which can be sustained in equilibrium. We assume that, whenever multiple equilibria exist, the Pareto dominant one is selected, as also done in Tabellini (2008b) and Attanasi et al. (2022).[20] In this case, the assumption about the selection criterion implies that in the region where $\theta^i \in [1, 2]$ the selected equilibrium strategy is (I, S). Therefore, the selected profiles of equilibrium strategies are:

- $\boldsymbol{s}^i = (O, T)$, if $\theta^i \leq 1$;
- $\boldsymbol{s}^i = (I, S)$, if $\theta^i > 1$.

Fig. 3 summarizes the behavior of individuals depending on the group of player B, and on the guilt sensitivity of the two populations. Notably, A's second order beliefs can be directly derived from equilibrium behavior. Specifically, $\beta_B^i(S) = 0$ when $\theta^i < 1$ and $\beta_B^i(S) = 1$ otherwise.

[17] We use Bayesian Sequential Equilibrium as it is the extension for psychological games of Sequential Equilibrium (Kreps and Wilson, 1982), which is a widely used refinement. We focus on BSE due to the need of selecting a unique equilibrium of the stage game for every parametric specification. See Battigalli et al. (2019a) for the definition of BSE, for an extensive discussion of solution concepts for psychological games and their properties, and for the characterization of BSE in pure strategies of the Trust Game with role-dependent guilt.

[18] In order to analyze the cultural transmission process we focus on pure strategy equilibria only.

[19] Note that we could extend our analysis to the case in which beliefs are incorrect using, for instance, selfconfirming equilibrium (refer to Battigalli, 1987; Fudenberg and Levine, 1993; Battigalli et al., 2015). In particular, we can show that (in addition to (O, T) and (I, S)) it may exist a selfconfirming equilibrium in which player A has wrong, but confirmed, first order beliefs about B's action and chooses to go Out even if B would be willing to Share. Note that this would enlarge the occurrences in which trust is not observed.

[20] See Appendix C of the working paper version of this paper for a discussion of the refinement criterion, and for the derivation of the result.

Table 1
Average material payoffs and psychological utilities, given the levels of guilt aversion and the group, when the matching is known.

Bounds of the region	Average material payoff/psychological utility	Equilibrium Strategies
$\theta^L \leq \theta^H < 1$	$\bar{m}_t^L = \bar{m}_t^H = 1 = \bar{u}_t^L = \bar{u}_t^H$	$s^H = s^L = (O,T)$
$\theta^L < 1 \leq \theta^H$	$\bar{m}_t^L = \frac{3-q_t}{2} = \bar{u}_t^L$, $\bar{m}_t^H = \frac{4-q_t}{2} = \bar{u}_t^H$	$s^L = (O,T)$, $s^H = (I,S)$
$1 \leq \theta^L \leq \theta^H$	$\bar{m}_t^L = \bar{m}_t^H = 2 = \bar{u}_t^L = \bar{u}_t^H$	$s^H = s^L = (I,S)$

Given that $\theta^L \leq \theta^H$, we have three relevant parametric regions for the analysis. For each of them, we characterize the equilibrium outcomes of the stage game given the match, at every time $t \in \mathbb{N}$.

Region 1 $\theta^L \leq \theta^H < 1$. In this region, for every type of match, there is only one sequential equilibrium outcome, that is, for each $i \in C$, $s^i = (O,T)$. At any interaction each individual plays in role A with probability $\frac{1}{2}$, in which case she meets a co-player of group L with probability q_t. However, regardless of the role and the group of the co-player, each player gains a material payoff of 1 in every interaction. Therefore, the average material payoffs at each time t are $\bar{m}_t^L = \bar{m}_t^H = 1$. In such a case, material payoffs and psychological utilities coincide.

Region 2 $\theta^L < 1 \leq \theta^H$. In this region the equilibrium strategies depend on the specific match. If the individual plays in role A (which happens with probability $\frac{1}{2}$) her payoff depends on the guilt sensitivity of the matched partner (and therefore on the composition of the population q_t). If he instead plays in role B, the equilibrium strategy and his payoff depend only on his guilt sensitivity. The average material payoffs are therefore

$$\bar{m}_t^L = \frac{1}{2}\left(q_t \cdot 1 + (1-q_t)\cdot 2\right) + \frac{1}{2}\cdot 1 = \frac{3-q_t}{2},$$
$$\bar{m}_t^H = \frac{1}{2}\left(q_t \cdot 1 + (1-q_t)\cdot 2\right) + \frac{1}{2}\cdot 2 = \frac{4-q_t}{2},$$

and they coincide with the average psychological utilities $\bar{u}_t^L$, $\bar{u}_t^H$.

Region 3 $1 \leq \theta^L \leq \theta^H$. In this region, for every type of match, there is only one sequential equilibrium outcome, that is, for each $i \in C$, $s^i = (I,S)$. At any interaction each individual plays in role A with probability $\frac{1}{2}$, in which case she meets a co-player of group L with probability q_t. However, regardless of the role and the guilt sensitivity of the co-player, each player gains a material payoff of 2 in every interaction. Therefore, the average material payoffs at each time t are $\bar{m}_t^L = \bar{m}_t^H = 2$. In such a case, material payoffs and psychological utilities coincide.

Note that, in all the three regions discussed above, as the path (I,T) never occurs, the average material payoffs coincide with the average psychological utilities (with both perfect and imperfect empathy) in each group, that is for each $i \in C$, $\bar{m}_t^i = \bar{u}_t^i = \hat{u}_t^i$, given that guilt is never experienced in equilibrium. Therefore, focusing on parental values, given that for each $i \in C$ $\mathbb{E}^i_{\delta_{q_t}}[q_{t+1}] = q_t$, we also have that $V_t^{ii} = V_t^{ji} = \bar{m}_t^i = \bar{u}_t^i = \hat{u}_t^i$. This means that the utility each parent expects to derive from a child of group i is independent of both her parental values and her group. Table 1 summarizes these results.

Let us now analyze how the results of Proposition 1 on cultural complementarity and substitution, hold when agents have complete information about the partner's group.

Corollary 1. *Under complete information about the matching, independently of the parenting style, the socialization efforts in equation* (5) *for both groups L and H always display cultural substitution.*

The result trivially follows from the functional form of the optimal socialization effort, $\tau_t^i = [(1-q_t^i)\Delta V_t^i]^+$. Indeed, as we can see from Table 1, for each $i \in C$, ΔV_t^i is independent of population shares, so that only the cultural substitution passing through $(1-q_t^i)$ is present.

4.1. Social dynamics

We now analyze the social dynamics. Let $q_\theta^* \in Q_\theta^* := \{q_t(\theta) \in [0,1] : \dot{q}_t(\theta) = 0\}$ denote a generic steady state of equation (6) at a given $\theta := (\theta^L, \theta^H)$, and s_θ^{i*} the corresponding steady-state equilibrium strategy for a generic $i \in C$. Note that we introduce a notation that highlights how the social dynamics may depend on the vector of guilt sensitivities θ.

Proposition 2. *Given the dynamics in equation* (6) *and complete information about the matching,* ***independently of the parenting style****,*

(i) *If $\theta^L \leq \theta^H < 1$, then $q_\theta^* = q_0$ and, for each $i \in C$, $s_\theta^{i*} = (O,T)$;*
(ii) *If $1 \leq \theta^L \leq \theta^H$, then $q_\theta^* = q_0$ and, for each $i \in C$, $s_\theta^{i*} = (I,S)$;*
(iii) *If $\theta^L < 1 \leq \theta^H$, then $Q_\theta^* = \{0,1\}$ and $q_\theta^* = 0$ is globally stable, and, for each $i \in C$, $s_\theta^{i*} = (I,S)$.*

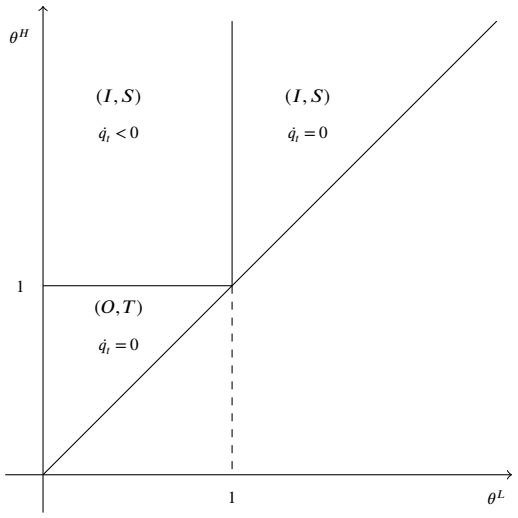


Fig. 4. Population dynamics and selected equilibrium strategies in the case of common knowledge of the matching.

Fig. 4 represents graphically the population dynamics and the steady-state equilibrium strategies observed described in Proposition 2.

Points (*i*) and (*ii*) of Proposition 2 state that, if the guilt sensitivities of the two groups are both high ($1 \leq \theta^L \leq \theta^H$) or both low ($\theta^L \leq \theta^H < 1$), any vector of population shares is a steady state as it delivers the same material payoffs and psychological utilities for agents of both groups. If $\theta^L \leq \theta^H < 1$ (Region 1), we always observe *O*ut as equilibrium path, independently of q_t. Indeed, each agent, when playing in role A, is aware of the fact that the partner, independently of his group, has such a low guilt sensitivity that he is not willing to *S*hare, thus any player A chooses *O*ut. Conversely, if $1 \leq \theta^L \leq \theta^H$ (Region 3), we always observe (I,S) as equilibrium path, independently of q_t. Indeed, the guilt sensitivity of both groups is so high that any player B always chooses *S*hare and this induces players A to go *I*n. In both cases, cultural heterogeneity persists in the long run. Note however that, this cultural heterogeneity is coupled with substantive homogeneity in behavior.

Point (*iii*) of Proposition 2 states that, when $\theta^L < 1 \leq \theta^H$ (Region 2), only agents of group H survive in the steady state and the observed equilibrium path is always (I,S). Given that player A has complete information about the guilt sensitivity of player B, she will go *I*n when matched with agents of group H, and *O*ut when matched with agents of group L. In the former case they both gain 1, whereas in the latter they gain 2. Therefore, the average payoff of agents who belong to group H is higher than the average payoff of agents who belong to group L. In this case, cultural heterogeneity is not sustained in the long run and agents of group H invade the society. It is interesting to notice that in this region the social dynamics leads to a homogeneous population in the long run, despite the presence of cultural substitution.[21]

Proposition 2 states that results are independent of parenting styles. Notably, even if parents are materialistic, guilt sensitivity plays a role for the evolution of the population dynamics. For example, consider a materialistic and selfish society where agents have no guilt aversion ($\theta^L = 0$) — i.e., the y-axis of Fig. 4. Suppose that a new small group of agents with positive guilt aversion ($\theta^H > 0$) enters in the society at time t so that $q_t = 1 - \varepsilon$, with an arbitrary $\varepsilon > 0$. If the guilt sensitivity of this minority group is small, $\theta^H < 1$, the trait is preserved in the long-run but has no effect on the outcome of the society. If, instead, the guilt sensitivity of the minority is large enough, $\theta^H > 1$, then, not only the trait is preserved but it also dominates in the long-run and leads the whole society to play (I,S). The result holds irrespectively of the size ε and of the parenting style.

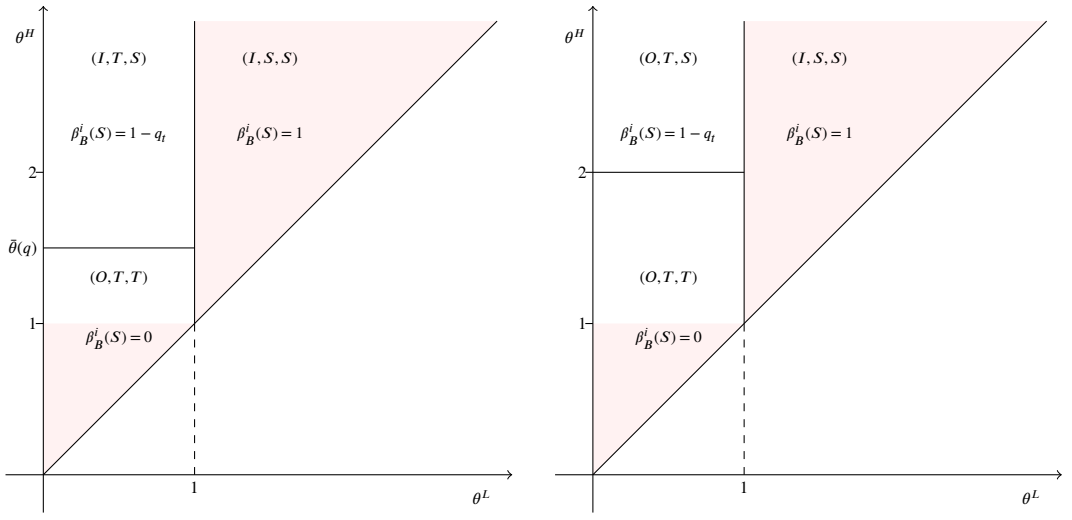
Overall, if at least one group has a high enough guilt sensitivity (i.e., $\theta^H \geq 1$) and agents are able to observe the guilt sensitivity of their partner, in the long run the level of trust of the society (namely the share of agents playing *I*n) reaches its maximum. The following remark presents the implications of the results above on the level of trust (worthiness).

Remark 1. When traits are observable, independently on the parenting style, the levels of trust and trustworthiness in the society weakly increase over time.

5. Incomplete information about the matching

The assumption of observability of the guilt aversion trait may hold in small communities, but is less realistic in large anonymous societies. Depending on the way agents interact in the society they live in, they may observe (or infer) the guilt aversion traits of their partners or not. As a matter of example, in large cities, in which interactions are much more sparse and happen daily also

[21] In this case, despite the fact that τ_t^L displays cultural substitution, the long-run homogeneity is reached because τ_t^H is always zero.



(a) $q \leq \frac{1}{2}$ (b) $q > \frac{1}{2}$

Fig. 5. Selected equilibrium behavior in the case of unknown matching. The (red-)shaded area represents the regions in which the selected equilibrium outcome is the same as in the complete information case.

among agents who do not know each other, it is very likely that the trait of the matched partner is not observed. Therefore, in this section, we study the long-run dynamics of guilt when agents in their adult age interact in a strategic environment where, for each interaction, there is incomplete information about the matching, i.e., player A does not know whether player B belongs to group H or L. As a consequence, player A forms beliefs on her expected payoff by combining her beliefs on the probability that players B from group $i \in C$ Share, with the information on the population shares, q_t. Note that the population shares, and in turn the strategies, now depend on t. However, for simplicity of notation, we drop the subscript t as long as we analyze the equilibrium of the Trust Game. We restore the dependence on t explicitly in Section 5.1 when we analyze the social dynamics. The disappointment of player A now depends on A's beliefs on both low guilt and high guilt B players, and on the population share of the low guilt group, q as follows:

$$D_A(s,\boldsymbol{\alpha}_A) = (1-\alpha_A^I)\cdot 1 + \alpha_A^I \cdot 2\left(q\alpha_A^{L,S} + (1-q)\alpha_A^{H,S}\right).$$

Note that the main difference between complete and incomplete information is that in the latter the agents' expected payoffs and, their strategies, depend on the population share q. In detail, the expected payoffs of players in role A depend on their beliefs about how many agents plan to play Share when in role B. This, in turn, affects the second-order beliefs of agents in role B, and their possible psychological loss.

A BSE in pure strategies is now constituted by a profile of strategies $\boldsymbol{s} = (s_A, s_B^L, s_B^H)$, together with a profile of (correct) second-order beliefs $\boldsymbol{\beta} = (\beta_A, \beta_B^L, \beta_B^H)$— where the first element of the vector refers to player A, the second to player B if belongs to group L, and the third to player B if belongs to group H. Note that, as far as player A is concerned, the equilibrium describes her behavior and her beliefs regardless of her group, as we assume that each player A holds the same (homogeneous) beliefs. Player B's behavior, instead, depends on whether he belongs to group L or H. Note that, the composition of the population q_t affects both player A's beliefs and her disappointment, and therefore it affects the equilibrium which is played.

Appendix B characterizes the (pure strategy) BSEs of this game. Some of the regions suffer of the same problem of multiplicity of equilibrium strategy profiles, as for the case with complete information. As before, we select equilibria according to the Pareto dominance criterion.[22]

Let us define $\bar{\theta}(q) := \frac{1}{1-q}$. Fig. 5 summarizes the selected equilibrium strategies and second order beliefs in different regions of the parameter space. To understand how incomplete information about the matching shapes behavior with respect to the complete information case, in Fig. 5 we highlight the areas in which the equilibrium outcomes are the same as in the complete information setting. Differently from the case of complete information, we report the second order beliefs because they are derived both from behavior and population shares. Thus, the dynamics of beliefs follows the dynamics of population shares.[23]

Let us now discuss the results keeping the same classification of regions as in Section 4 for ease of comparison.

Region 1 ($\theta^L \leq \theta^H < 1$) Agents always choose Out when in role A and Take when in role B, regardless of their group, as in the complete information case.

[22] Further details of the equilibrium selection procedure can be found in Appendix C of the working paper version of the paper.

[23] Note that second order beliefs of B players from different groups coincide, given that A does not know B's group and therefore A's first order belief does not depend on B's identity.

Table 2
Average material payoffs and psychological utilities, given the levels of guilt aversion, the group, and the population share, when the matching is unknown.

Pop. share	Bounds of the region	Average material payoff	Average psychological utility
$q_t \leq \frac{1}{2}$	$\theta^L < 1$ and $\theta^H < \bar{\theta}(q_t)$	$\bar{m}^L = \bar{m}^H = 1$	$\bar{u}^L = \bar{u}^H = 1$
	$\theta^L < 1$ and $\theta^H \geq \bar{\theta}(q_t)$	$\bar{m}^L_t = 3 - q_t$, $\bar{m}^H_t = 2 - q_t$	$\bar{u}^L_t = 3 - q_t - \theta^L(1 - q_t)$, $\bar{u}^H_t = 2 - q_t$
	$\theta^L \geq 1$ and $\theta^H > 1$	$\bar{m}^L = \bar{m}^H = 2$	$\bar{u}^L = \bar{u}^H = 2$
$q_t > \frac{1}{2}$	$\theta^L < 1$	$\bar{m}^L = \bar{m}^H = 1$	$\bar{u}^L = \bar{u}^H = 1$
	$1 \leq \theta^L \leq \theta^H$	$\bar{m}^L = \bar{m}^H = 2$	$\bar{u}^L = \bar{u}^H = 2$

Region 2 ($\theta^L < 1 \leq \theta^H$) Let us first consider players in role A. Note that L agents have such a low guilt sensitivity that, when in role B, they always Take, independently of the population shares. Therefore, if $q > \frac{1}{2}$ (Fig. 6b), independently of what agents H choose, an agent in role A always prefers to go Out. Conversely, if $q < \frac{1}{2}$ (Fig. 6a), the optimal choice depends on what agents H do when playing in role B: if they Share, the optimal choice for players in role A is to go In, so that the equilibrium strategy is (I, T, S); if they Take, the optimal choice is to go Out and the equilibrium strategy is (O, T, T).

Consider now agents in role B. As we discussed above, agents L always Take. Instead, the optimal choices of agents H when in role B depend on both q and their own guilt sensitivity. The threshold $\bar{\theta}(q)$ is the guilt sensitivity for which, given q, agents H are indifferent between Sharing and Taking. For higher levels of guilt sensitivity ($\theta^H > \bar{\theta}(q)$), agents H always Share, and for lower levels ($\theta^H < \bar{\theta}(q)$) agents H always Take.

Note that, if q increases, the area in which the equilibrium strategy profile is (O, T, T) grows larger, given that $\bar{\theta}(q) = \frac{1}{1-q}$ is increasing in q. As a matter of fact, when q_t is large, the expected payoff of player A from going In is low and so is her disappointment.

Region 3 ($1 \leq \theta^L \leq \theta^H$) In this region, the equilibrium strategy profile that survives the Pareto dominance selection is (I, S, S). Thus, the equilibrium outcome is the same as with complete information for every possible matching.

Table 2 reports the equilibrium average material payoffs and psychological utilities in the different parametric regions. Note that, psychological utilities differ from material payoffs only when the selected equilibrium prescribes (I, T, S).

As we did in the previous section, we now analyze the results about cultural complementarity and substitution of Proposition 1 when agents have incomplete information about the partner's group.

Corollary 2. *Under incomplete information about the matching,*

- τ^H_t *displays cultural complementarity if and only if parents are imperfectly empathic and* $\theta^L < 1$ *and* $\theta^H \geq \bar{\theta}(q_t)$*;*
- τ^L_t *displays cultural complementarity if and only if parents are (perfectly or imperfectly) empathic and* $\frac{1}{2(1-q_t)} < \theta^L < 1$ *and* $\theta^H \geq \bar{\theta}(q_t)$*.*

From Corollary 2 we see that socialization efforts may display cultural complementarity only when the equilibrium strategy profile is (I, T, S) — i.e., in the area $\theta^L < 1$ and $\theta^H \geq \bar{\theta}(q_t)$ — and only if parents take into account children's psychological utility when choosing the optimal socialization effort. Indeed, from Table 2 it is straightforward to see that only in this parameter space the cultural intolerance may depend on q_t.

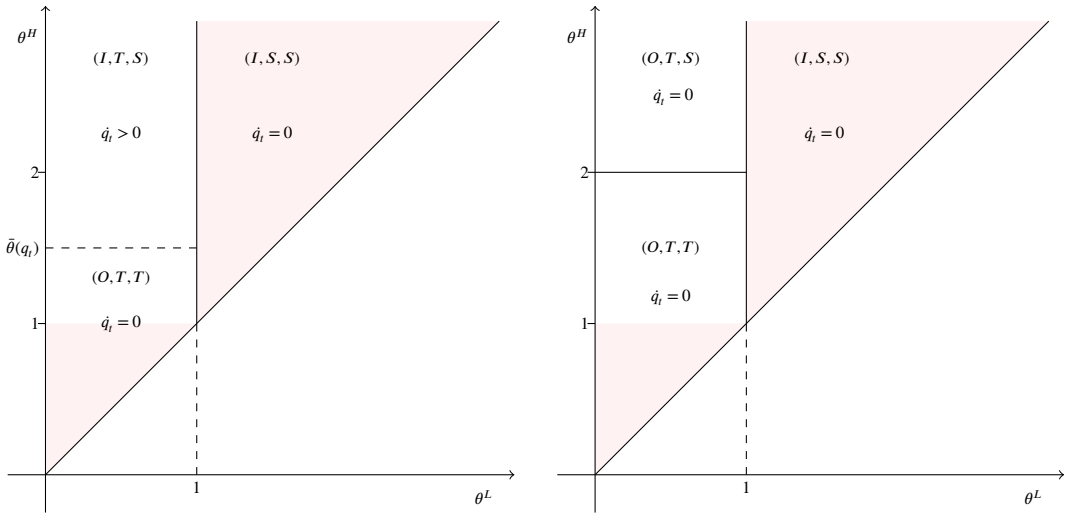
In this area, when parents are imperfectly emphatic the optimal socialization of parents of group H always displays cultural complementarity because the higher their share, the higher the share of matches in which the cooperative path (I, S) is played and the higher the payoffs.[24]

Consider low guilt agents. As shown Table 2, L agents always have a material advantage over H agents, but the $\bar{m}^L_t - \bar{m}^H_t$ is independent of q_t. Thus, if parents were just materialistic, only cultural substitution would have been displayed. If parents have positive guilt sensitivity and are empathic, their guilt towards H agents would counterbalance this effect —i.e., a higher share of L agents in the society implies that they feel less guilty and also that they have higher incentives to transmit their own trait. However, for low guilt sensitivity levels (i.e., $\theta^L < \frac{1}{2(1-q_t)}$) this effect is not strong enough and we always see substitution. On the contrary, if guilt sensitivity is higher (i.e., $\theta^L > \frac{1}{2(1-q_t)}$) complementarity arises.

5.1. Social dynamics

We now consider the population dynamics induced by the equilibrium strategies. Given that material and psychological payoffs differ for some parameter space (as shown in Table 2), we have that the dynamics may differ depending on the parenting styles. Indeed, by looking at equation (6), it is clear that the dynamics is characterized by the difference in the socialization efforts, which in turn depends on the way parents evaluate children's expected payoffs, i.e., on their parenting styles, as discussed in Section 2.3.

[24] Notably, if they were perfectly empathic and, thus, evaluated the future psychological payoffs of low guilt children with the correct psychological parameter, they would have never socialized them because $\Delta V^H < 0$.



(a) $q_t \leq \frac{1}{2}$ (b) $q_t > \frac{1}{2}$

Fig. 6. Population dynamics and selected equilibrium strategies in the case unknown matching and materialistic or perfectly empathic parents. The (red-)shaded area represents the regions in which the selected equilibrium outcome is the same as in the complete information case.

In what follows, we present the full characterization of the population dynamics and the steady-state strategies for the whole parameter space θ and for any possible q_0 (see Figs. 6 and 7 and Table 2).

We begin by characterizing the dynamics in the regions where they are independent from parenting styles (Proposition 3). We then focus on the most interesting case, where different parenting styles generate different dynamics (Propositions 4 and 5).

Proposition 3. *Consider the dynamics in equation* (6) *with incomplete information about the matching, and fix* (θ, q_0). *If either (i)* $q_0 > \frac{1}{2}$, *or (ii)* $q_0 \leq \frac{1}{2}$, $\theta^L < 1$, *and* $\theta^H \leq \bar{\theta}(q_0)$, *or (iii)* $q_0 \leq \frac{1}{2}$ *and* $1 \leq \theta^L < \theta^H$, *then,* ***independently of parenting style,*** $\dot{q}_t(\theta) = 0$, *so that* $q^*_\theta = q_0$.

The proposition characterizes the dynamics in those regions in which the average psychological utility is the same across groups and it coincides with the average material payoff, as there is no psychological loss from guilt (see Table 2). This happens in regions in which players in role A go Out and in those where all agents behave alike regardless of their group —as both guilt sensitivities are very high or very low. In all these cases, independently of the parenting styles, $\dot{q}_t(\theta) = 0$ and the steady-state population share coincides with q_0.

Let us now focus on those cases in which the parenting style plays a role, namely when $q_0 \leq \frac{1}{2}$, $\theta^L < 1$, and $\theta^H \geq \bar{\theta}(q_0)$. Note that the threshold $\bar{\theta}(q_t)$ evolves together with the population share, and this must be taken into account for the characterization of the steady state q^*_θ.

Materialism and perfect empathy If parents are materialistic, only material payoffs play a role in determining the optimal socialization efforts. In such a case, the material advantage in favor of group L, induced by the equilibrium strategy (I,T,S), makes $\tau^L > \tau^H = 0$, so that the share of agents of group L increases. The same holds in the presence of perfect empathy, as the material advantage that agents of group L have from Taking in role B is larger than their psychological loss from guilt. As a matter of fact, agents belonging to group L could ensure themselves the same utility as agents of group H by Sharing in role B, but they have a higher psychological utility from Taking. Proposition 4 characterizes the dynamics and the steady state for this case. Let $\bar{q}_{\theta^H} := \{q : \theta^H = \bar{\theta}(q)\}$.

Proposition 4. *Consider the dynamics in equation* (6) *with* (M) *or* (PE), *incomplete information about the matching, and fix* (θ, q_0), *with* $q_0 < \frac{1}{2}$. *If* $\theta^L < 1$ *and* $\theta^H \geq \bar{\theta}(q_0)$, *then* $\dot{q}_t(\theta) \geq 0$, $q^*_\theta = \min\{\bar{q}_{\theta^H}, \frac{1}{2}\}$, *and* $s^*_\theta = (I,T,S)$.

From Proposition 4, we can see that, if parents are materialistic or perfectly empathic, in the considered region the share of group L in the society, q_t, increases. When this happens, $\bar{\theta}(q_t)$ increases as well up to the point where either $\bar{\theta}(q_t) = \theta^H$ or $q_t = \frac{1}{2}$.

The Fig. 6 summarizes the social dynamics in the presence of materialistic and perfectly empathic parents, merging the results of Proposition 3 and 4. Specifically, in Fig. 6b we see that whenever $q_t > \frac{1}{2}$ there is no dynamics, whereas Fig. 6a shows that when $q_t \leq \frac{1}{2}$ there is a region in which the share of agents of group L in society increases.

Let us now compare the social dynamics with incomplete information with the one with complete information (i.e., Fig. 4). We find a difference in Region 2 ($\theta^L < 1 \leq \theta^H$). In particular, if $q_t > \frac{1}{2}$ (Fig. 6b), there is no dynamics and the population shares remain fixed independently on the initial conditions. On the other hand, if $q_t \leq \frac{1}{2}$, when $\theta^H \in [1, \bar{\theta}(q))$ all the agents gain the same utility and population shares are fixed over time, given that player A always goes Out; whereas, if $\theta^H \geq \bar{\theta}(q)$ the share of agents of group

L increases. Therefore, we can conclude that if parents are either materialistic or perfectly empathic, incomplete information favors agents with low guilt sensitivity, L, and always guarantees cultural heterogeneity in the long-run.

The implication of the dynamics on the level of trust and trustworthiness are contained in the following remark.

Remark 2. When traits are not observable, if parents are materialist or perfectly empathic the level of trustworthiness in the society weakly decreases over time.

Imperfect empathy Let us now consider parents who evaluate the psychological utilities of children using their own guilt sensitivity. Let us recall that, in the region where $\theta^L < 1$, $\theta^H \geq \bar{\theta}(q_0)$, and $q_0 \leq \frac{1}{2}$, the psychological utility of low guilt agents is $\bar{u}_t^L = 3 - q_t - \theta^L(1-q_t)$, whereas high guilt agents do not face any psychological loss and, thus, $\bar{u}^H = \bar{m}^H = 2 - q_t$ (see Table 2). Under imperfect empathy high guilt parents evaluate the psychological utility of agents belonging to group L as $\hat{u}_t^L = 3 - q_t - \theta^H(1-q_t)$. Therefore, given $\theta^H > \theta^L$, high guilt parents overestimate the eventual psychological loss of a child of group L and, thus, they exert a higher socialization effort with respect to perfectly empathic parents. Low guilt parents, on the contrary, exert the same socialization effort τ^L as perfectly empathic parents, because on the one hand they evaluate their children's psychological utility with the correct θ^L, and, on the other hand, their evaluation of group H utility is not affected by any assumption about parents' empathy because they do not experience any guilt. Let us define the following values: $\bar{\theta}'(q_t) := \frac{1-(1-q_t)^2\theta^L}{q_t(1-q_t)}$, $\hat{\theta}(q_t) := \frac{1-2q_t}{(1-q_t)^2}$, $\bar{q}_{\theta^H} := \{q : \theta^H = \bar{\theta}'(q_0)\}$, and $\hat{q}_{\theta^L} := \{q : \theta^L = \hat{\theta}(q_0)\}$.[25]

Proposition 5. *Consider the dynamics in equation (6) with (IE) and incomplete information about the matching. Fix (θ, q_0) with $q_0 < \frac{1}{2}$. If $\theta^L < 1$ and $\theta^H \geq \bar{\theta}(q_0)$, then,*

- *if $\theta^H < \bar{\theta}'(q_0)$, then $\dot{q}_t \geq 0$. Moreover, if $\theta^L < \hat{\theta}(q_0)$ then $q_\theta^* = \min\{\bar{q}_{\theta^H}, \bar{q}'_{\theta^H}, \hat{q}_{\theta^L}\}$, whereas, if $\theta^L \geq \hat{\theta}(q_0)$ then $q_\theta^* = \min\{\frac{1}{2}, \bar{q}_{\theta^H}\}$;*
- *If $\theta^H \geq \bar{\theta}'(q_0)$, then $\dot{q}_t \leq 0$. Moreover, if $\theta^L \leq \hat{\theta}(q_0)$ then $q_\theta^* = \hat{q}_{\theta^L}$, whereas, if $\theta^L > \hat{\theta}(q_0)$ then $q_\theta^* = \bar{q}_{\theta^H}$.*

Proposition 5 shows that imperfect empathy mitigates the positive effect of incomplete information on q_t, allowing the share of L agents to decrease in the society and, thus, allowing the overall level of guilt sensitivity to increase. Notably, this happens only if group H agents have a high enough guilt sensitivity. In this case, H agents, when evaluating the psychological loss of a child of group L, use their high guilt sensitivity parameter and overestimate the eventual psychological loss of a child of group L. For this reason, parents of group H have a high incentive to directly socialize children to own trait, so that their share in the society increases, i.e., q_t decreases. We can see in Fig. 7 that, under imperfect empathy, if $q_0 \leq \frac{1}{2}$ there is the area $\theta^H \geq \bar{\theta}'(q_0)$ in which $\dot{q}_t < 0$. The case in which $q_0 > \frac{1}{2}$ is described by Fig. 6b as the parenting style does not matter when $q_0 > \frac{1}{2}$.

The implication of the dynamics on the evolution of trust and trustworthiness in the presence of imperfect empathy are summarized by the following remark

Remark 3. When traits are not observable, if parents are imperfectly empathic the level of trustworthiness in the society weakly decrease over time if $\theta^H < \bar{\theta}'(q_0)$ and weakly increases if $\theta^H \geq \bar{\theta}'(q_0)$.

6. The role of homophily

So far we have assumed that the matching of agents in the society was random and independent of players' type. However, in reality this seldom happens. Indeed, it is a well-known fact in the literature that agents are more prone to interact with people with similar traits (Currarini et al., 2009). This phenomenon is known as *homophily*.

Let $a \in [0,1]$ be the *inbreeding* homophily rate which biases the random matching.[26] Let us also define ρ_t^i as the probability, at time t, of an agent of group $i \in C$ to meet an agent of the same group. Note that, generically, $\rho^i \neq 1 - \rho^j$. Specifically,

$$\begin{cases} \rho_t^L = a + (1-a)q_t \\ \rho_t^H = a + (1-a)(1-q_t). \end{cases}$$

When we introduce homophily, also the strategy of player A might depend on the group she belongs to, as her expectation on B's strategy (correctly) depends on her group. Consider for example an equilibrium in which B players from group L *T*ake, and B players from group H *S*hare. In the presence of homophily, A players from group H have a higher probability of being matched (and therefore a higher α_A^S) than A players from group L. As a consequence, a strategy profile has now length 4.

[25] Note that both $\bar{\theta}'(q_t)$ and $\hat{\theta}(q_t)$ are larger than $\bar{\theta}(q_t)$.

[26] We call the homophily rate a, as homophily induces *a*ssortative matching between agents.

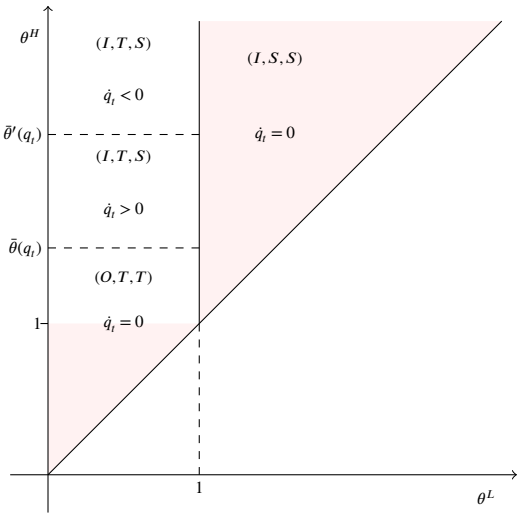


Fig. 7. Population dynamics and selected equilibrium strategies in the case of unknown matching, imperfectly emphatic parents and $q_0 \leq \frac{1}{2}$. The (red-)shaded area represents the regions in which the selected equilibrium outcome is the same as in the complete information case.

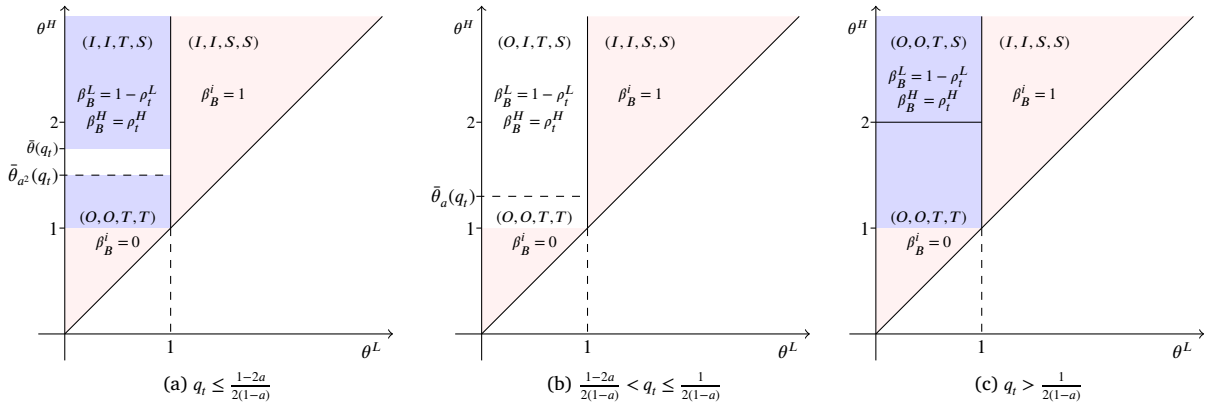


Fig. 8. Selected equilibrium strategies with unknown matching and homophily. The (red-)light-shaded area represents the regions in which the selected equilibrium outcome is the same as in the complete information and incomplete information without homophily cases. The (blue-)dark-shaded area represents the regions in which the selected equilibrium outcome is the same as in the incomplete information case without homophily (but different from the complete information case).

Specifically, player A's probability of being matched with a player B from group L now depends on the group of player A. We denote these probabilities with $\hat{q}_t^k$, where $k = H, L$ denotes the group player A belongs to. These probabilities are $\hat{q}_t^L = \rho^L$ and $\hat{q}_t^H = 1 - \rho^H$, respectively.

As a consequence, A's disappointment depends on A's group as follows:

$$D_A(s, \boldsymbol{\alpha}_A) = (1 - \alpha_A^{i,I}) \cdot 1 + \alpha_A^{i,I} \cdot 2\left(\hat{q}^i \alpha_A^{L,S} + \left(1 - \hat{q}^i\right) \alpha_A^{H,S}\right).$$

Note also that player B's belief on A's belief on being matched with a B player from group L now depends on his group.

Let us define, the thresholds $\bar{\theta}_a(q_t) := \frac{1}{1-(1-a)q_t}$ and $\bar{\theta}_{a^2}(q_t) := \frac{1}{1-(1-a^2)q_t}$, which coincide with $\bar{\theta}_a(q_t)$ when there is no homophily (i.e., $a = 0$).

Fig. 8 shows the selected equilibrium strategy profiles and second order beliefs with homophily. Note that, in this case, second order beliefs may differ between groups, given that matching probabilities, and therefore A's first order beliefs, are group dependent. The characterization of the equilibria and the selection procedure are contained in Appendix B.3.

We can see in the figure below the two thresholds on q_t that delimits the three panels (i.e., $\frac{1}{2(1-a)}$ and $\frac{1-2a}{2(1-a)}$) are decreasing a increasing in a, respectively. This, implies that as the level of assortativity increases, the space of q_t for which the equilibrium strategies are described by Panel b increases.

Table 3 reports the equilibrium average material payoffs in the different parametric regions in the case with homophily.

Table 3
Average material payoffs given the levels of guilt aversion, the group, and the population share, when the matching is unknown and there is homophily.

Pop. share	Bounds of the region	Average material payoff
$q_t \leq \frac{1-2a}{2(1-a)}$	$\theta^L < 1$ and $\theta^H < \bar{\theta}_{a^2}(q_t)$	$\bar{m}^L = \bar{m}^H = 1$
	$\theta^L < 1$ and $\theta^H \geq \bar{\theta}_{a^2}(q_t)$	$\bar{m}_t^L = 3 - \rho_t^L$, $\bar{m}_t^H = 1 + \rho_t^H$
	$\theta^L \geq 1$ and $\theta^H > 1$	$\bar{m}^L = \bar{m}^H = 2$
$\frac{1-2a}{2(1-a)} < q_t \leq \frac{1}{2(1-a)}$	$\theta^L < 1$ and $\theta^H < \bar{\theta}_a(q_t)$	$\bar{m}^L = \bar{m}^H = 1$
	$\theta^L < 1$ and $\theta^H \geq \bar{\theta}_a(q_t)$	$\bar{m}_t^L = \frac{5}{2} - \frac{3}{2}\rho_t^L$, $\bar{m}_t^H = \frac{1}{2} + \frac{3}{2}\rho_t^H$
	$\theta^L \geq 1$ and $\theta^H > 1$	$\bar{m}^L = \bar{m}^H = 2$
$q_t > \frac{1}{2(1-a)}$	$\theta^L < 1$	$\bar{m}^L = \bar{m}^H = 1$
	$1 \leq \theta^L \leq \theta^H$	$\bar{m}^L = \bar{m}^H = 2$

Lastly note that, since we are considering just materialistic parents, and the cultural intolerance with materialistic parents is always independent on q_t, the socialization efforts always display cultural substitution.

Social dynamics In what follows we analyze the impact of homophily on the population dynamics, in the presence of materialistic parents.

Let us start noticing that, whenever the equilibrium strategies are (O,O,T,T), (O,O,T,S), or (I,I,S,S), the payoffs are the same for both groups so that the population shares show no dynamics. Therefore, by looking at Fig. 8, if $q_0 > \frac{1}{2(1-a)}$ or if $\theta^L > 1$ or $\theta^H < \bar{\theta}_a(q_0)$, then $\dot{q} = 0$.[27]

Proposition 6 provides the main insights from the dynamics for the areas in which population shares do change. Note that a full characterization of steady states can be found in the proof of the proposition in Appendix A.

Proposition 6. *Consider the dynamics (6) with (M) and homophily, at $\theta^L < 1$, $\theta^H \geq \bar{\theta}_a(q_0)$, $q_0 \leq \frac{1}{2(1-a)}$. Then,*

- *If $a \leq 1/3$, then $\dot{q}_t \geq 0$ and the steady state is weakly increasing in θ^H;*
- *If $a > 1/3$, then:*
  - *if $q_0 > \frac{1-2a}{2(1-a)}$, then $\dot{q}_t < 0$;*
  - *if $q_0 \leq \frac{1-2a}{2(1-a)}$, then $\dot{q}_t > 0$ if $\theta^H \geq \bar{\theta}_{a^2}\left(\frac{1-2a}{2(1-a)}\right)$, whereas $\dot{q}_t = 0$ otherwise.*

From Proposition 6, we can see that, if homophily is low (i.e., $a < \frac{1}{3}$), then agents of group L, interacting often enough with agents of group H, can exploit them and, thus, get advantage of partners' high guilt, so that $\dot{q}_t > 0$. unsurprisingly, this result is in line with what happens for the case without homophily. Moreover, the higher θ^H, the more L agents are better-off and, thus, their share in the society at the steady state is larger.

As homophily increases, low guilt agents have a lower possibility to exploit high guilt agents, as they are less often matched with them.

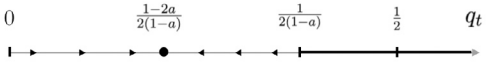
If homophily is high enough (i.e., $a > \frac{1}{3}$) agents of both groups interact among themselves with a large enough frequency, so that their payoffs are mostly affected by the equilibrium strategies played by the agents belonging to their own group. In particular, when the share of agents of group L is sufficiently high in the society (i.e., $q_0 > \frac{1-2a}{2(1-a)}$), low guilt agents interact mainly among themselves and their low guilt makes them worse-off with respect to H agents. On the contrary, if the share of L agents is low (i.e., $q_0 \leq \frac{1-2a}{2(1-a)}$), then, even for relatively high levels of homophily, they interact with many H agents. Therefore, if the guilt parameter of H agents, θ^H, is above a certain threshold, L agents can exploit it and gain higher payoffs, hence $\dot{q}_t > 0$. If instead the guilt parameter of H agents is below the threshold, agents of both groups choose Out since the level of trustworthiness is so low that no one trusts any other agent (and thus $\dot{q}_t = 0$).

The implication of the population dynamics on the level of trust and trustworthiness are summarized in the following remark (see also Fig. 9).

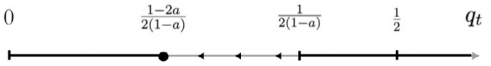
Remark 4. Let us consider non-observable traits, homophily, and materialistic parents.

- If homophily is low ($a \leq 1/3$), the levels of trust and trustworthiness in the society weakly decrease over time.

[27] Note that $\dot{q}_t = 0$ even when $q_t \leq \frac{1-2a}{2(1-a)}$, $\theta^L < 1$, and $\bar{\theta}_{a^2}(q_t) \leq \theta^H < \bar{\theta}_a(q_t)$. This case is internalized in Proposition 6.



(a) for $\theta^L < 1$, $\theta^H \geq \bar{\theta}_a(q_0)$, when $\theta^H \geq \bar{\theta}_{a^2}\left(\frac{1-2a}{2(1-a)}\right)$.



(b) for $\theta^L < 1$, $\theta^H \geq \bar{\theta}_a(q_0)$, when $\theta^H < \bar{\theta}_{a^2}\left(\frac{1-2a}{2(1-a)}\right)$.

Fig. 9. Social Dynamics with homophily, when $a > \frac{1}{3}$.

- If homophily is high ($a > 1/3$), the levels of trust and trustworthiness increase for high population shares whereas for low population shares the levels of trust and trustworthiness weakly decrease.

7. Discussion

Alternative modeling assumptions We model guilt aversion as a role-dependent psychological trait that activates only when the individual plays in role B. One may wonder how the results would change if players were motivated by guilt aversion also when finding themselves in role A (*role-independent guilt aversion*). The Trust Game, with role-independent guilt aversion, has been analyzed in Attanasi et al. (2016) for the case of incomplete information. Allowing player A to be guilt averse may increase her propension to choose In, if B expects her to do it. In other words, A may find it optimal to go In even for first order beliefs on B's sharing lower than $\frac{1}{2}$. This has two effects: (i) it changes the bounds of the regions for existence of some equilibria when we have incomplete information on the matching—for example, the equilibrium (I, O, S, T) may now exist even for some $q_t > \frac{1}{2}$—; (ii) it introduces the possibility of equilibria in which A chooses to go In even when she attaches little to no probability to player B choosing to Share. In these equilibria, B correctly believes that A will go In and would be disappointed by her choice of staying Out. To avoid this disappointment, A goes In if her guilt sensitivity is high enough to overcome the loss in material payoff associated with the path (I, T). In these equilibria, guilt aversion increases trust in absence of trustworthiness. Societies in which individuals feel compelled to trust others even when they expect them not to be trustworthy are difficult to observe. As a matter of fact, even in the theoretical model, if we consider a society in which high guilt players choose In in role A and Take in role B while low guilt players choose Out and Take, we will observe the population share of the high guilt group decreasing steadily until the high-guilt population disappears.

Another interesting extension of the model, mentioned in Section 4 above, is to adopt *selfconfirming equilibrium* (SCE) (Battigalli, 1987; Fudenberg and Levine, 1993; Battigalli et al., 2015) instead of BSE, as a solution concept. SCE, differently from BSE, allows agents to have incorrect beliefs that are nevertheless confirmed on the equilibrium path. With SCE, we can have equilibria in which there is little or no trust, even in presence of trustworthiness—the opposite of what we observed with role-independent guilt aversion. For example, (O, S) can be the equilibrium strategy profile. To understand how this is possible, imagine if player A has a pessimistic belief about the behavior of B— i.e., she expects that he would act selfishly if given the opportunity. Given her (incorrect) beliefs, A stays Out even if B would be willing to Share. By staying Out, A does not have the opportunity to realize that her beliefs are incorrect and, therefore, (O, S) can be the equilibrium strategy profile.

Guilt and religion As discussed in the Introduction, guilt and trust are strongly related to religious beliefs and values, particularly for some religions (Weber, 2013; Guiso et al., 2003; Walinga et al., 2005; Sheldon, 2006; Oviedo, 2016). Thus, our model speaks to the literature that investigates the link between the evolution of religious beliefs and the evolution of trust (e.g., Shariff and Norenzayan, 2007; Norenzayan and Shariff, 2008; Norenzayan, 2013) and can be easily read as an analysis of the conditions under which a religion survives and expands in a population.

Under this interpretation, the model predicts that, on the one hand, in small closed societies where everyone knows each other and, thus, the commitment to religious values is observable, we should expect greater levels of religiosity than in large anonymous societies, where the commitment to religious values is mostly not observable. This finding is supported by the evidence that people who live in rural areas display greater religiosity than their urban counterparts (e.g., Chalfant and Heller, 1991). Furthermore, also Norenzayan and Shariff (2008) suggest that the association of religiosity—and actual prosocial behavior and trust—emerges primarily in contexts where reputations concerns are heightened and there is an apparent profession of religious devotion.

On the other hand, the paper confirms the intuition that religiosity is essential for supporting cooperation in large anonymous societies (Shariff and Norenzayan, 2007; Norenzayan and Shariff, 2008; Norenzayan, 2013). In fact, with incomplete information, what we observe is that while the religious trait never completely dominates, it can thrive only if parents have imperfect empathy,

which, as discussed earlier, can originate from the fear of some moralizing god. Thus, the paper aligns with the theory that beliefs in moralizing gods play a crucial role in fostering cooperation in large societies.

Additionally, the literature summarized by Norenzayan et al. (2016) argues that religiosity and belief in moralizing gods are actually prerequisites for the emergence of large, anonymous societies. While our paper does not directly address this specific issue, it provides supporting evidences for this theory. In fact, in our model, consistent with the evidence (Durkheim and Swain, 1915; Louch, 2000; Cheadle and Schwadel, 2012) that religious groups tend to self-segregate partially or fully,[28] we demonstrate that in anonymous societies the religious group finds it profitable to self-segregate. This can be observed by noting that, for high levels of homophily, the material payoffs of the high-guilt group are higher or equal to the ones without homophily.[29] As a result, if agents were to have the choice to self-segregate, the society would fragment into two smaller societies where the individuals' types become common knowledge, no longer constituting a large anonymous society. Examining situations in which the profit of self-segregation decreases for the high guilt group, we clearly observe that this happens either when everyone has a sufficiently high guilt—i.e., high religiosity—to play (I, S); or if the share of low-guilt type is small—i.e., the difference in payoffs between the cases with and without homophily is decreasing in q—; or when there is imperfect empathy—e.g., moralizing gods—which ensures that the share of individuals with high guilt increases in society. These findings support the idea that religiosity and moralizing gods are essential factors in sustaining cooperation and the existence of large anonymous societies.

8. Conclusion

This is the first paper that studies the intergenerational transmission of a psychological trait (guilt aversion) that affects strategic decisions of agents with belief-dependent preferences and the derived consequences in terms of long-run evolution of cooperation, trust, and trustworthiness.

Agents are the more cooperative the higher their guilt sensitivity, as the desire to avoid guilt feelings induces them to share in the Trust Game. When agents' guilt sensitivity is observable, socialization efforts always display cultural substitution and the share of agents with low guilt weakly decreases over time, so that, trust and trustworthiness increase. Conversely, if agents' guilt sensitivity is not observable, the material advantage of high guilt agents disappears and, in most cases, the share of the low guilt population weakly increases, benefiting from betraying the trust of their partners who cannot recognize them as low guilt. In such a case, the share of high guilt agents can weakly increase only if parents are imperfectly empathic or if there is homophily in society.

In this paper we focused on the belief-dependent preference that we believe is the main drive of cooperation in Trust Game (see Attanasi et al., 2019, 2022, for supporting evidence). However, social and belief-dependent preferences are heterogeneous across individuals, and motivations like—e.g., fairness concerns and reciprocity may coexist with guilt aversion, or be the prevalent motivation for some individuals. Exploring the dynamics of population shares when groups have fundamentally different cultures—i.e., are motivated by social preferences of different nature—is an interesting topic for future research.

Additionally, the paper suggests that institutions may affect the evolution of trust, guilt, and religions themselves, which in turn may change the institutional setting itself. We strongly believe that such a relation between cultural transmission of psychological traits and the evolution of institutions is a topic that deserves to be investigated further.

Declaration of competing interest

Elena Manzoni: None. **Fabrizio Panebianco:** None.

Data availability

No data was used for the research described in the article.

Acknowledgments

Sebastiano Della Lena gratefully acknowledges fund from FWO-foundation for the project "Diffusion of Misinformation in Social Networks" (id. 42933).

Appendix A. Proofs of propositions

A.1. Proof of Proposition 1

Let us consider two generic group i and j. The socialization effort of i, τ^i, displays cultural substitution if and only if:

$$\frac{\partial \tau_t^i}{\partial q_t^i} = -\Delta V_t^i + (1 - q_t^i)\frac{\partial \Delta V_t^i}{\partial q_t^i} < 0$$

[28] We refer to Razin (2019) for a deep discussion about religion and segregation.

[29] See Table 2 and Table 3, and Appendix A.

$$\Rightarrow \quad -\frac{\partial \Delta V_t^i}{\partial q_t^j} \underbrace{\frac{q_t^j}{\Delta V_t^i}}_{\geq 0} < 1$$

The result follows from the last inequality. ■

A.2. Proof of Proposition 2

In Regions 1 and 3, for every $i, j \in C$, $V_t^{ij} = V_t^{ii}$ so that $\Delta V_t^i = \Delta V_t^j = 0$. Then, given equation (5), independently of q_t, $\tau_t^{i*} = \tau_t^{j*} = 0$, so that any q_t is a steady state.
Consider now Region 2. By construction of (6), $q_t = 0$ and $q_t = 1$ are always steady states. Moreover, $\bar{m}_t^L = \frac{3-q_t}{2} < 2 - \frac{q_t}{2} = \bar{m}_t^H$, and consequently $0 > \Delta V_t^L = -\Delta V_t^H$. Then, for every q_t, $\tau_t^{H*} > \tau_t^{L*} = 0$ and $\dot{q}_t < 0$ for every $q_t \in (0,1)$. Consequently $q_t = 0$ is globally stable. Equilibrium actions follows. ■

A.3. Proof of Corollary 2

We can see from Table 2 that, for materialistic parents and all the parameter spaces different from $\theta^L < 1$ and $\theta^H \geq \bar{\theta}(q_t)$, cultural intolerances ΔV^L and ΔV^H do not depend on the population shares. Therefore, using the result of Proposition 1, socialization effort always displays cultural substitution. Let us now focus on empathic parents and $\theta^L < 1$ and $\theta^H \geq \bar{\theta}(q_t)$. For low guilt parents, substituting the value of ΔV_t^L as in Table 2, there is cultural complementarity if and only if:

$$\frac{\partial \tau_t^L}{\partial q_t} = -\Delta V_t^L + (1-q_t)\frac{\partial \Delta V_t^L}{\partial q_t} > 0$$

$$\Rightarrow \quad \theta^L > \frac{1}{2(1-q_t)}$$

Let us consider high guilt imperfectly empathic parents, substituting the value of ΔV_t^H as in Table 2, there is cultural complementarity if and only if:

$$\frac{\partial \tau_t^H}{\partial (1-q_t)} = -\Delta V_t^H + q_t \frac{\partial \Delta V_t^H}{\partial (1-q_t)} > 0$$

$$\Rightarrow \quad \theta^H > 1 \quad \text{always.}$$

If instead high guilt parents are perfectly emphatic $\Delta V_t^H = 1 - \theta^L(1-q_t)$, which with $\theta^L < 1$ is always negative, thus, $\tau_t^H = 0$. ■

A.4. Proof of Proposition 3

Consider the payoffs agents get when $\theta^L < \bar{\theta}(q_t)$, and $\theta^H \geq \bar{\theta}(q_t)$ provided in Table 2 and in Appendix B. In all these regions $m^H = m^L = u^H = u^L$. Then, independently of the parenting style $\dot{q} = 0$. ■

A.5. Proof Proposition 4

Materialistic parents In this region, the equilibrium is (I,T,S). Recall also that $\theta^H \geq \bar{\theta}(q_t)$ implies $q_t \leq \frac{\theta^H - 1}{\theta^H}$, and that $\theta^L < \bar{\theta}(q_t)$ implies $q_t > \frac{\theta^L - 1}{\theta^L}$. As shown in Table 2 and in Appendix B, $m^L = 3 - q_t$ and $m^H = 2 - q_t$. Then, for each $q_t \in (\frac{\theta^L - 1}{\theta^L}, \frac{\theta^H - 1}{\theta^H}]$, $m^L > m^H$. Because parents are materialistic, $\tau^{L*} > \tau^{H*}$. Thus $\dot{q}_t > 0$ and $q_\theta^* = min\{\frac{\theta^H - 1}{\theta^H}, \frac{1}{2}\}$.

Perfectly empathic parents In this region, the equilibrium is (I,T,S), therefore

$$\begin{cases} \bar{u}^L &= \frac{1}{2} \cdot \left[(1-q_t)\cdot 2 + q_t \cdot 0\right] + \frac{1}{2} \cdot \left[4 - \theta^L \cdot 2(1-q_t)\right] = 3 - q_t - \theta^L(1-q_t) \\ \bar{u}^H &= \frac{1}{2} \cdot \left[(1-q_t)\cdot 2 + q_t \cdot 0\right] + \frac{1}{2} \cdot 2 \ = 2 - q_t. \end{cases}$$

Note also that $\bar{u}^L > \bar{u}^H$ and, thus, $\tau^L > \tau^H$ if and only if $\theta^L < \frac{1}{1-q_t}$ which always holds in this region. Thus, we have that $\tau^L > \tau^H$ always and thus $\dot{q}_t > 0$ and $q_\theta^* = min\{\frac{\theta^H - 1}{\theta^H}, \frac{1}{2}\}$. ■

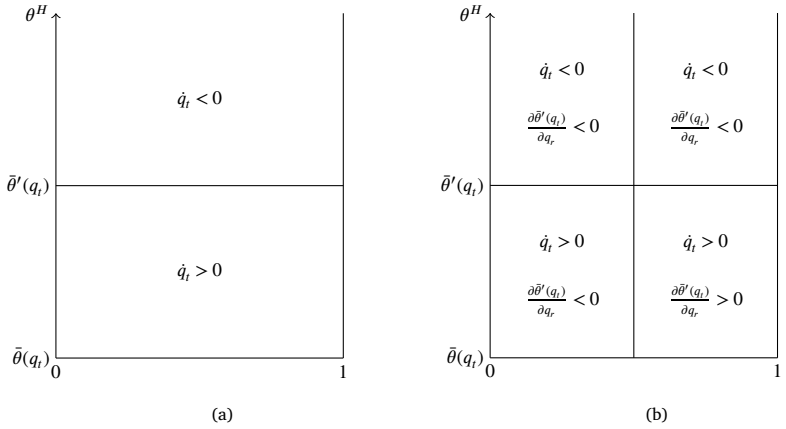


Fig. 10. Population dynamics in the case of role-dependent guilt, unknown matching, and imperfectly emphatic parents, in the space $\theta \in [0,1] \times [\bar{\theta}(q_t), \infty]$.

A.6. Proof Proposition 5

Recall that the equilibrium in this region is (I,T,S). Then, the psychological utilities are

$$\begin{cases} \bar{u}^L & = \frac{1}{2} \cdot \left[(1-q_t) \cdot 2 + q_t \cdot 0\right] + \frac{1}{2} \cdot \left[4 - \theta^L \cdot 2(1-q_t)\right] = 3 - q_t - \theta^L(1-q_t), \\ \hat{u}^L & = \frac{1}{2} \cdot \left[(1-q_t) \cdot 2 + q_t \cdot 0\right] + \frac{1}{2} \cdot \left[4 - \theta^H \cdot 2(1-q_t)\right] = 3 - q_t - \theta^H(1-q_t), \\ \bar{u}^H = \hat{u}^H & = \frac{1}{2} \cdot \left[(1-q_t) \cdot 2 + q_t \cdot 0\right] + \frac{1}{2} \cdot 2 = 2 - q_t. \end{cases}$$

Consider first low guilt agents. Note that $\bar{u}^L > \hat{u}^H$ if and only if $4 - 2\theta^L(1-q_t) > 2$, that is, $\theta^L < \frac{1}{1-q_t} = \bar{\theta}(q_t)$. Thus, since $\theta^L < 1$, we always have that $\tau^{L*} > 0$.

Consider now high guilt agents. Note that $\bar{u}^H > \hat{u}^L$ if and only if $4 - 2\theta^H(1-q_t) < 2$, that is, $\theta^H > \frac{1}{1-q_t}$. This is always the case because, in this area, $\theta^H \geq \bar{\theta}(q_t)$. Then, $\tau^{H*} > 0$ always.

Consider first τ^{L*}.

$$\begin{aligned} \tau^{L*} &= \frac{1}{2}(1-q_t)\Big[4 - 2\theta^L(1-q_t) - 2\Big] \\ &= (1-q_t)\Big(1 - (1-q_t)\theta^L\Big). \end{aligned}$$

Consider now τ^{H*}

$$\begin{aligned} \tau^{H*} &= \frac{1}{2}q_t\Big[2 - \big(4 - 2\theta^H(1-q_t)\big)\Big] \\ &= q_t\Big(-1 + (1-q_t)\theta^H\Big). \end{aligned}$$

The social dynamics is determined by the difference in parental efforts, which is given by the following:

$$\tau^{L*} - \tau^{H*} = (1-q_t)\Big(1 - (1-q_t)\theta^L\Big) - q_t\Big(-1 + (1-q_t)\theta^H\Big).$$

Therefore,

$$\tau^{L*} - \tau^{H*} > 0 \quad \textit{if and only if} \quad \theta^H < \frac{1 - (1-q_t)^2\theta^L}{q_t(1-q_t)} =: \bar{\theta}'(q_t).$$

Notice that $\bar{\theta}'(q_t) > \bar{\theta}(q_t)$ always. Therefore, in the Region 2, $[0,1] \times [\bar{\theta}(q_t), +\infty]$, where the PBE is (I,T,S), the dynamics is such that if $\theta \in [0,1] \times \big[\bar{\theta}(q_t), \bar{\theta}'(q_t)\big)$, then $\dot{q}_t > 0$; whereas, if $\theta \in [0,1] \times [\bar{\theta}'(q_t), \infty]$, then $\dot{q}_t < 0$. This is reported in Fig. 10a below.

To study the dynamics given a pair (θ^L, θ^H) we first need to analyze how $\bar{\theta}'(q_t)$ changes with q_t.

$$\frac{\partial \bar{\theta}'(q_t)}{\partial q_t} = \frac{\theta^L + \theta^L q_t^2 - 2\theta^L q_t + 2q_t - 1}{(1-q_t)^2 q_t^2} = \frac{\theta^L(1-q_t)^2 + 2q_t - 1}{(1-q_t)^2 q_t^2}.$$

Then, we have that

$$\frac{\partial \bar{\theta}'(q_t)}{\partial q_t} > 0 \quad \textit{if and only if} \quad \theta^L > \frac{1 - 2q_t}{(1-q_t)^2} =: \hat{\theta}(q_t).$$

Moreover $\frac{\partial\hat{\theta}(q_t)}{\partial q_t} < 0$ always. Note finally that whenever $\theta^H = \bar{\theta}'(q_t)$, $\theta^H = \bar{\theta}(q_t)$, or $\theta^L = \hat{\theta}(q_t)$, then $\dot{q}_t = 0$. Therefore, considering the space in Region 2, Fig. 10b represents the four areas derived by the thresholds $\hat{\theta}(q_t)$ and $\bar{\theta}'(q_t)$.

To analyze the dynamics and characterize the steady state, let us define the following:

$$\tilde{q}^* := \{q : \theta^H = \bar{\theta}'(q), 0 < q < \min\{\frac{1}{2}, \frac{\theta^H - 1}{\theta^H}\}, 0 < \theta^L < 1\} = \frac{\theta^H - 2\theta^L + \sqrt{(\theta^H)^2 + 4\theta^L - 4\theta^H}}{2(\theta^H - \theta^L)} > 0$$

$$\hat{q}^* := \{q : \theta^L = \hat{\theta}(q), 0 < q < \min\{\frac{1}{2}, \frac{\theta^H - 1}{\theta^H}\}, \theta^L < 1\} = 1 - \frac{1 - \sqrt{1 - \theta^L}}{\theta^L} \in \left[0, \frac{1}{2}\right].$$

Let us consider in turn the areas of the Fig. 10b.

- If $\theta^H \geq \bar{\theta}'(q_t)$ and $\theta^L \leq \hat{\theta}(q_t)$ (i.e., $q_t \leq \hat{q}^*_\theta \leq \bar{q}'_{\theta^H\theta}$), then $\dot{q}_t \leq 0$. Thus, as q_t decreases, $\bar{\theta}'(q_t)$ increases, $\hat{\theta}(q_t)$ increases, and $\bar{\theta}(q_t)$ decreases. Given a point (θ^L, θ^H) in this area, and given how $\bar{\theta}(q_t)$, $\bar{\theta}'(q_t)$, $\hat{\theta}(q_t)$ move with q_t, then the dynamics stops when $\theta^H = \bar{\theta}'(q_t)$, so that $q^*_\theta = \bar{q}'_{\theta^H}$.
- if $\theta^H \geq \bar{\theta}'(q_t)$ and $\theta^L > \hat{\theta}(q_t)$ (i.e., $q_t \geq \bar{q}'_{\theta^H\theta}$ and $q_t \geq \hat{q}^*_\theta$), $\dot{q}_t \leq 0$. Thus, as q_t decreases, $\bar{\theta}'(q_t)$ decreases, $\hat{\theta}(q_t)$ increases, and $\bar{\theta}(q_t)$ decreases. Given a point (θ^L, θ^H) in this area, and given how $\bar{\theta}(q_t)$, $\bar{\theta}'(q_t)$, $\hat{\theta}(q_t)$ move with q_t, then the dynamics stops when $\theta^L = \hat{\theta}(q_t)$, so that $q^*_\theta = \hat{q}^*$.
- if $\theta^H < \bar{\theta}'(q_t)$ and $\theta^L \geq \hat{\theta}(q_t)$ (i.e., $\hat{q}^*_\theta \leq q_t \leq \bar{q}'_{\theta^H\theta}$), $\dot{q}_t \geq 0$. Thus, as q_t increases, $\bar{\theta}'(q_t)$ increases, $\hat{\theta}(q_t)$ decreases, and $\bar{\theta}(q_t)$ increases. Given a point (θ^L, θ^H) in this area, and given how $\bar{\theta}(q_t)$, $\bar{\theta}'(q_t)$, $\hat{\theta}(q_t)$ move with q_t, then the dynamics stops when $\theta^H = \bar{\theta}(q_t)$ or if q_t reaches $\frac{1}{2}$, so that $q^*_\theta = \begin{cases} \frac{\theta^H - 1}{\theta^H} & \text{if} \quad \theta^H < 2 \\ \frac{1}{2} & \text{if} \quad 2 < \theta^H < 4 - \theta^L \end{cases}$, so that $q^*_\theta = \min\{\frac{1}{2}, \frac{\theta^H - 1}{\theta^H}\}$.
- if $\theta^H < \bar{\theta}'(q_t)$ and $\theta^L < \hat{\theta}(q_t)$ (i.e., $q_t \leq \bar{q}'_{\theta^H\theta}$ and $q_t \leq \hat{q}^*_\theta$), $\dot{q}_t \geq 0$. Thus, as q_t increases, $\bar{\theta}'(q_t)$ decreases, $\hat{\theta}(q_t)$ decreases, and $\bar{\theta}(q_t)$ increases. Given a point (θ^L, θ^H) in this area, and given how $\bar{\theta}(q_t)$, $\bar{\theta}'(q_t)$, $\hat{\theta}(q_t)$ move with q_t, then all the threshold may be binding, thus, the dynamics stops when $q^*_\theta = \min\{\frac{\theta^H - 1}{\theta^H}, \tilde{q}^*, \hat{q}^*\}$. ■

A.7. Proof Proposition 6

To prove results in Proposition 6, we present and prove two auxiliary propositions that delivers a complete characterization of the dynamics and steady states. Proposition 6 presents the main insights from these two auxiliary propositions.

Proposition 7. *Consider the dynamics in (6) with homophily, at $\theta^L < 1$, $\theta^H > \bar{\theta}_a(q_0)$, $q_0 \leq \frac{1}{2(1-a)}$, and $a \leq \frac{1}{3}$. Then $\dot{q} \geq 0$. Moreover,*

- *If $q_0 \leq \frac{1-2a}{2(1-a)}$ and*
  - *$\theta^H \in \left[\bar{\theta}_a(q_0), \bar{\theta}_{a^2}(q_0)\right)$, then $q^* = q_0$;*
  - *$\theta^H \in \left[\bar{\theta}_{a^2}(q_0), \bar{\theta}_{a^2}(\frac{1-2a}{2(1-a)})\right)$, then $q^* = \frac{\theta^H - 1}{\theta^H(1-a^2)}$;*
  - *$\theta^H \in \left[\bar{\theta}_{a^2}(\frac{1-2a}{2(1-a)}), 2\right)$, then $q^* = \frac{\theta^H - 1}{\theta^H(1-a)}$;*
  - *$\theta^H \geq 2$, then $q^* = \frac{1}{2(1-a)}$;*
- *If $q_0 \in \left(\frac{1-2a}{2(1-a)}, \frac{1}{2(1-a)}\right]$ and*
  - *$\theta^H \in (\bar{\theta}_a(q_0), 2)$, then $q^* = \frac{\theta^H - 1}{\theta^H(1-a)}$;*
  - *$\theta^H > 2$, then $q^* = \frac{1}{2(1-a)}$.*

Proposition 8. *Consider the dynamics (6) with homophily, at $\theta^L < 1$, $\theta^H \geq \bar{\theta}_a(q_0)$, when $q_0 \leq \frac{1}{2(1-a)}$ and $a > \frac{1}{3}$. Then,*

- *if $\theta^H \geq \bar{\theta}_{a^2}(\frac{1-2a}{2(1-a)})$, then $q^* = \frac{1-2a}{2(1-a)}$ is the globally stable steady state;*
- *if $\theta^H \in \left[\bar{\theta}_a(q_0), \bar{\theta}_{a^2}(\frac{1-2a}{2(1-a)})\right)$ then $q^* = \begin{cases} q_0 & \text{if} \quad q_0 \leq \frac{1-2a}{2(1-a)}; \\ \frac{1-2a}{2(1-a)} & \text{if} \quad q_0 > \frac{1-2a}{2(1-a)}. \end{cases}$*

Proof of Proposition 7 and 8 Before proceeding with the proofs let us notice that:

$$\begin{cases} \bar{\theta}_a\left(\frac{1}{2(1-a)}\right) = \frac{1}{1-(1-a)\frac{1}{2(1-a)}} = 2; \\ \bar{\theta}_{a^2}\left(\frac{1-2a}{2(1-a)}\right) = \frac{2}{2-(1+a)(1-2a)}. \end{cases}$$

Let us consider the spaces where the equilibria are either (O,I,T,S) —i.e., $\frac{1-2a}{2(1-a)} < q_0 \leq \frac{1}{2(1-a)}$, $\theta^L < 1$, $\theta^H \geq \bar{\theta}_a(q_0)$ — or (I,I,T,S) —i.e., $q_0 \leq \frac{1-2a}{2(1-a)}$, $\theta^L < 1$, $\theta^H \geq \bar{\theta}_{a^2}(q_0)$.

– (O,I,T,S) In such a case the **material payoffs** are:

$$\begin{cases} \bar{m}^L = & \frac{1}{2}\cdot 1 + \frac{1}{2}\left(\rho^L \cdot 1 + (1-\rho^L)\cdot 4\right) = \frac{5}{2} - \frac{3}{2}\rho^L; \\ \bar{m}^H = & \frac{1}{2}\cdot\left(\rho^H \cdot 2 + (1-\rho^H)\cdot 0\right) + \frac{1}{2}\left(\rho^H \cdot 2 + (1-\rho^H)\cdot 1\right) = \frac{1}{2} + \frac{3}{2}\rho^H. \end{cases}$$

Let us compute the difference

$$\bar{m}^L - \bar{m}^H = 2 - \frac{3}{2}(\rho^L + \rho^H) = 2 - \frac{3}{2}(1+a) = \frac{1}{2}(1-3a),$$

so that $\bar{m}^L > \bar{m}^H$ if and only if $a < \frac{1}{3}$. Note also that defining $\Delta\bar{m}^L$ and $\Delta\bar{m}^H$ the differences between the material payoffs in the case with and without homophily for L and H respectively, we can see that, if $\theta^H > \bar{\theta}(q_0)$, $\Delta\bar{m}^L = -\frac{1}{2}(1+q+3a(1-q)) < 0$ always and $\Delta\bar{m}^H = -\frac{1}{2}(1-3a)q > 0$ if and only if $a > \frac{1}{3}$. We can also see that in the cases in which $\Delta\bar{m}^H > 0$ the larger the size of H—i.e., the smaller q—the lower is the increment in profits stemming from self-segregation. If instead $\theta^H \in [\bar{\theta}_a(q_0), \bar{\theta}(q_0)]$ $\Delta\bar{m}^L > 0$ always.

– (I,I,T,S) In such a case the **material payoffs** are:

$$\begin{cases} \bar{m}^L = & \frac{1}{2}\cdot\left((1-\rho^L)\cdot 2 + \rho^L \cdot 0\right) + \frac{1}{2}\cdot 4 = 3 - \rho^L = 3 - a - (1-a)q; \\ \bar{m}^H = & \frac{1}{2}\cdot\left(\rho^H \cdot 2 + (1-\rho^H)\cdot 0\right) + \frac{1}{2}\cdot 2 = \rho^H + 1 = 1 + a + (1-a)(1-q) = 2 - q + aq. \end{cases}$$

Thus, $\bar{m}^L > \bar{m}^H$ and $\dot{q}_t > 0$ always. Note also that, if $\theta^H > \bar{\theta}(q_0)$ $\Delta\bar{m}^L = -a + aq < 0$ and $\Delta\bar{m}^H = aq > 0$. Namely, High-guilt agents find more profitable interacting a society with homophily, whereas it is less profitable for Low-guilt agents. Moreover, both $\Delta\bar{m}^L$ and $\Delta\bar{m}^H$ are decreasing in the size of group H—i.e. increasing in q. If instead $\theta^H \in [\bar{\theta}^2_a(q_0), \bar{\theta}(q_0)]$ $\Delta\bar{m}^L > 0$ always.

If $a = \frac{1}{3}$ In this case, if $q_0 \in \left(\frac{1-2a}{2(1-a)}, \frac{1}{2(1-a)}\right]$ where (O,I,T,S) are the equilibrium strategies, then $\bar{m}^L = \bar{m}^H$ and, thus, $\dot{q}_t = 0$. If instead $q_0 \leq \frac{1-2a}{2(1-a)}$ where (I,I,T,S) are the equilibrium strategies and, thus, $\dot{q}_t \geq 0$.

If $a < \frac{1}{3}$

- Let us consider the case of $q_0 \in \left(\frac{1-2a}{2(1-a)}, \frac{1}{2(1-a)}\right]$, where (O,I,T,S) are the equilibrium strategies (see Fig. 8b). In such cases, $\dot{q}_t > 0$. Note that $\frac{\partial\bar{\theta}_a(q_t)}{\partial q_t} > 0$. Starting from any point in the space (θ, q_0) where (O,I,T,S) are the equilibrium strategies, either q_t increases (and thus $\bar{\theta}_a(q_t)$ does so) up to the point that $\theta^H = \bar{\theta}_a(q_t)$, or $\bar{\theta}_a(q_t)$ never reaches θ^H so that q_t keeps increasing until $q_t = \frac{1}{2(1-a)}$, where either (O,O,T,T) or (O,O,T,S) are the equilibrium strategies, and the dynamics stops. Therefore, the steady state is either $q^* = \{q : \theta^H = \bar{\theta}_a(q)\} = \frac{\theta^H - 1}{\theta^H(1-a)}$, or $q^* = \frac{1}{2(1-a)}$. The steady state is $q^* = \min\{\frac{\theta^H-1}{\theta^H(1-a)}, \frac{1}{2(1-a)}\}$. Note also that $\frac{\theta^H-1}{\theta^H(1-a)} \geq \frac{1}{2(1-a)}$ if and only if $\theta^H \geq 2$ and that $\bar{\theta}_a(q_t = \frac{1}{2(1-a)}) = 2$. Thus, $q^* = \begin{cases} \frac{\theta^H-1}{\theta^H(1-a)} & \text{if} \quad \theta^H \leq 2 \\ \frac{1}{2(1-a)} & \text{if} \quad \theta^H > 2 \end{cases}$.
- Let us consider the case of $q_0 \in \left[0, \frac{1-2a}{2(1-a)}\right]$ (see Fig. 8a). It is trivial to see that if $\theta^H \in (\bar{\theta}_a(q_0), \bar{\theta}_{a^2}(q_0))$ then (O,O,T,T) are the equilibrium strategies and, thus, $\dot{q} = 0$ so that $q^* = q_0$. Let us now consider $\theta^H \geq \bar{\theta}_{a^2}(q_0)$ where (I,I,T,S) are the equilibrium strategies and note that $\frac{\partial\bar{\theta}_{a^2}(q_t)}{\partial q_t} > 0$. In such a case $\dot{q}_t > 0$ and either q_t increases until $q_t = \frac{1-2a}{2(1-a)}$ and (O,I,T,S) are the equilibrium strategies and the analysis in the previous bullet point holds, or $\bar{\theta}_{a^2}(q_t)$ increases until it reaches θ^H and (O,O,T,T) are the equilibrium strategies and the steady state is $q^* = \{q : \theta^H = \bar{\theta}_{a^2}(q)\} = \frac{\theta^H-1}{\theta^H(1-a^2)}$. Therefore, if θ^H is higher than $\bar{\theta}_{a^2}(q_0)$ but lower than the upper-bound of the threshold — i.e., $\bar{\theta}_{a^2}(\frac{1-2a}{2(1-a)})$ — the dynamics stops when $\bar{\theta}_{a^2}(q_t) = \theta^H$, so that $q^* = \frac{\theta^H-1}{\theta^H(1-a^2)}$. If instead θ^H is higher than $\bar{\theta}_{a^2}(\frac{1-2a}{2(1-a)})$, q_t increases and, at a some t, it overcomes $\frac{\theta^H-1}{\theta^H(1-a^2)}$ where the equilibrium strategies become (O,I,T,S) and the analysis of the previous bullet point holds. Thus, if $\theta^H \geq 2$ then $q^* = \frac{1}{2(1-a)}$, whereas if $\theta^H \in [\bar{\theta}_{a^2}(\frac{1-2a}{2(1-a)}), 2)$, then $q^* = \frac{\theta^H-1}{\theta^H(1-a)}$.

If $a > \frac{1}{3}$

- Let us consider the regions where (O,I,T,S) (i.e., $\frac{1-2a}{2(1-a)} < q_0 \leq \frac{1}{2(1-a)}$) are the equilibrium strategies. In such cases $\dot{q}_t < 0$ and q_t decreases over time. Thus,

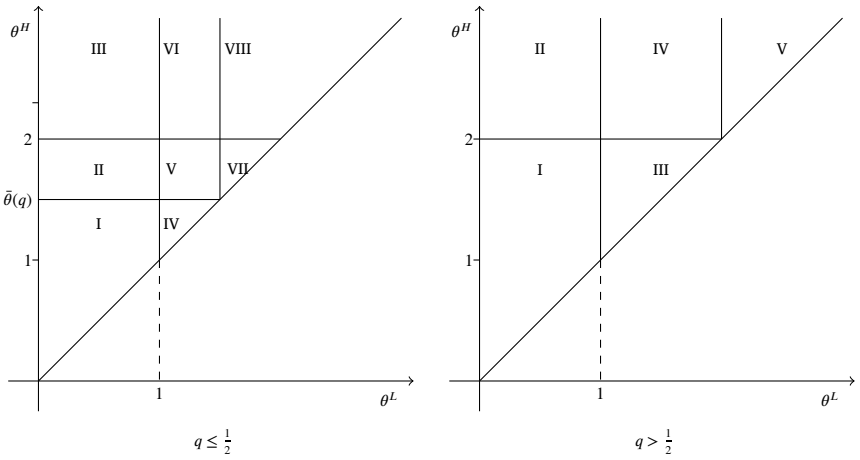


Fig. 11. Regions of analysis, unknown matching.

- If $\theta^H \geq 2$, then as soon as $q_t = \frac{1-2a}{2(1-a)}$, then (I, I, T, S) are the equilibrium strategies and, as previously argued, q_t should be increasing. Thus, $q^* = \frac{1-2a}{2(1-a)}$ is the steady state.
- If $2 > \theta^H \geq \bar{\theta}_{a^2}(q_t = \frac{1-2a}{2(1-a)})$ then, again, as soon as $q_t = \frac{1-2a}{2(1-a)}$, then (I, I, T, S) are the equilibrium strategies and, as previously argued, q_t should be increasing. Thus, $q^* = \frac{1-2a}{2(1-a)}$ is the steady state.
- If $\theta^H < \bar{\theta}_{a^2}(q_t = \frac{1-2a}{2(1-a)})$, then as $q_t = \frac{1-2a}{2(1-a)}$ the dynamics stops.

• Let us consider the regions where (I, I, T, S) (i.e., $q_0 \leq \frac{1-2a}{2(1-a)}$) are the equilibrium strategies. In this case $\dot{q}_t > 0$ and q_t increases over time. Thus,

- If $\theta^H \geq 2$, then as soon as $q_t = \frac{1-2a}{2(1-a)}$, then (O, I, T, S) should be played and, as previously argued, q_t should start to decrease. Thus, $q^* = \frac{1-2a}{2(1-a)}$ is the steady state.
- If $2 > \theta^H \geq \bar{\theta}_{a^2}\left(q_t = \frac{1-2a}{2(1-a)}\right)$ then, again, as soon as $q_t = \frac{1-2a}{2(1-a)}$, then (O, I, T, S) are the equilibrium strategies and, as previously argued, q_t should start to decrease. Thus, $q^* = \frac{1-2a}{2(1-a)}$ is the steady state.
- If $\theta^H < \bar{\theta}_{a^2}\left(q_t = \frac{1-2a}{2(1-a)}\right)$, then as $q_t = \frac{1-2a}{2(1-a)}$, then q_t keeps increasing up to the point that $q^* = \frac{\theta^H - 1}{\theta^H(1-a^2)}$, where $\theta^H = \bar{\theta}_{a^2}(q_t)$ and (O, O, T, T) are the equilibrium strategies. ■

Proposition 6 trivially follows from Proposition 7 and 8.

Appendix B. Characterization of the equilibria with incomplete information of the matching

We now propose the characterization of equilibria in pure strategies under incomplete information over the match. Equilibria are analyzed separately for the case where $q < \frac{1}{2}$ and the case where $q > \frac{1}{2}$. For the formal definition of BSE in this context we refer to Battigalli et al. (2019a). Note that our game is *naive*, in the definition of Battigalli et al. (2019a), in that there is only one epistemic type for each player, i.e., $\mathcal{E}_i$ is a singleton, hence we do not introduce formally epistemic types.

B.1. Case I: $q < \frac{1}{2}$

The left panel of Fig. 11 highlights the regions that we consider.

B.1.1. Region I: $\theta^L < 1$ and $\theta^H < \bar{\theta}(q)$

B's best response is to Take if he has low guilt θ^L, independently of his second-order belief β_B^L. Hence, A's first-order belief $\alpha_A^S \leq 1 - q$, and so is $\beta_B^i(S)$, $i \in \{L, H\}$. If A chose to enter $\alpha_A^S \geq \frac{1}{2}$, hence her disappointment after (I, T), $D_A(I, T, \alpha_A) = 1 - \alpha_A^I + 2\alpha_A^S\alpha_A^I$ is increasing both in α_A^S and in α_A^I. The maximum disappointment (and the maximum guilt for player B) is reached when $\alpha_A^I = 1$ and $\alpha_A^S = 1 - q$, i.e., A expects B to Share if his guilt sensitivity is high. If $\theta^H < \bar{\theta}(q) = \frac{1}{1-q}$, B's utility from Taking is higher than B's utility from Sharing even when his guilt is high:

$$4 - \theta^H(1 - \alpha_A^I + 2\alpha_A^S\alpha_A^I) > 4 - \frac{1}{1-q}(2(1-q)) \geq 2.$$

Therefore the pure equilibrium strategies are (O, T, T), and the second-order beliefs are the degenerate ones that obtain from the equilibrium strategies.

B.1.2. Region II: $\theta^L < 1$ and $\bar{\theta}(q) \leq \theta^H < 2$

B's best response is to Take if he has low guilt sensitivity θ^L, independently of his second-order belief β_B^L. As in Region 1, A's first-order belief $\alpha_A^S \leq 1-q$, and so is $\beta_B^i(S)$, $i \in \{L, H\}$. Now, two equilibria may arise, depending on the second-order beliefs of the high guilt player B, as θ^H is sufficiently high to sustain an equilibrium in which he Shares:

(i) (O,T,T) with $\alpha_A^I = 0$, $\alpha_A^S = 0$, and, by correct conjectures, $\mathbb{E}[D_A, \beta_B | I] = 1 < \frac{2}{\theta^H}$ given $\theta^H < 2$.
(ii) (I,T,S) with $\alpha_A^I = 1$, $\alpha_A^S = 1-q$, and $\mathbb{E}[D_A, \beta_B | I] = 2(1-q) \geq \frac{2}{\theta^H}$, given $\theta^H \geq \frac{1}{1-q}$. A finds it optimal to go In even when only high guilt B players Share because $q < \frac{1}{2}$.

B.1.3. Region III: $\theta^L < 1$ and $\theta^H \geq 2$

B's best response is to Take if he has low guilt sensitivity θ^L, and to Share if he has high guilt sensitivity, independently of his second-order belief. Given that the fraction of low guilt B players is $q < \frac{1}{2}$ player A finds it optimal to go In even when only high guilt B players Share. Therefore the pure equilibrium strategies are (I,T,S), and the second-order beliefs are the degenerate ones that obtain from the equilibrium strategies.

B.1.4. Region IV: $1 \leq \theta^L < \bar{\theta}(q)$ and $\theta^H < \bar{\theta}(q)$

Both the low and the high guilt B may find it optimal to $Share$ or $Take$, depending on A's beliefs. As we focus on pure strategy equilibria, they can both Share, both Take, or it can be that the high guilt B Shares (Takes) and the low guilt B Takes (Shares) respectively. A's best response is to go In if at least the high guilt B Shares, and to stay Out if at most the low guilt B does it. Two of these strategy profiles are equilibria:

(i) (O,T,T) with $\alpha_A^I = 0$, $\alpha_A^S = 0$, and, by correct conjectures, $\mathbb{E}[D_A, \beta_B | I] = 1 < \frac{2}{\theta^i}$, for $i \in C$, given $\theta^L < 2$ and $\theta^H < 2$.
(ii) (I,S,S) with $\alpha_A^I = 1$, $\alpha_A^S = 1$ and by correct conjectures, $\mathbb{E}[D_A, \beta_B | I] = 2 > \frac{2}{\theta^i}$, for $i \in C$, given $\theta^L > 1$ and $\theta^H > \frac{1}{1-q}$.

The other two strategy profiles are not equilibria: (I,T,S) induces $\mathbb{E}[D_A, \beta_B | I] = 2(1-q) < \frac{2}{\theta^H}$, given $\theta^H < \frac{1}{1-q}$; (O,S,T) induces $\mathbb{E}[D_A, \beta_B | I] = 2q < \frac{2}{\theta^L}$, given $\theta^L > 1$ and $q < \frac{1}{2}$.

B.1.5. Region V: $1 \leq \theta^L < \bar{\theta}(q)$ and $\bar{\theta}(q) \leq \theta^H < 2$

Both the low and the high guilt B player may optimally Share or Take, depending on A's beliefs. As in Region IV, they can both Share, both Take, or that one Share and one Take. A's best response is to go In if at least the high guilt B Shares, and to stay Out if at most the low guilt B does it. Three of these strategy profiles are equilibria:

(i) (O,T,T) with $\alpha_A^I = 0$, $\alpha_A^S = 0$, and, by correct conjectures, $\mathbb{E}[D_A, \beta_B | I] = 1 < \frac{2}{\theta}$ given $\theta^L < 2$ and $\theta^H < 2$.
(ii) (I,T,S) with $\alpha_A^I = 1$, $\alpha_A^S = 1-q$, and $\mathbb{E}[D_A, \beta_B | I] = 2(1-q) > \frac{2}{\theta^H}$, given $\theta^H > \frac{1}{1-q}$. Note that $\mathbb{E}[D_A, \beta_B | I] = 2(1-q) < \frac{2}{\theta^L}$.
(iii) (I,S,S) with $\alpha_A^I = 1$, $\alpha_A^S = 1$ and by correct conjectures, $\mathbb{E}[D_A, \beta_B | I] = 2 > \frac{2}{\theta}$ given $\theta^L > 1$ and $\theta^H > \frac{1}{1-q}$.

The other strategy profile is not an equilibrium one: (O,S,T) induces $\mathbb{E}[D_A, \beta_B | I] = 2q < \frac{2}{\theta^L}$, given $\theta^L > 1$ and $q < \frac{1}{2}$.

B.1.6. Region VI: $1 \leq \theta^L < \bar{\theta}(q)$ and $\theta^H \geq 2$

B's best response is to Share if he has high guilt sensitivity. Hence, A always goes In, regardless of the choice of low guilt B, because $q < \frac{1}{2}$. We have two possible equilibria:

(i) (I,T,S) with $\alpha_A^I = 1$, $\alpha_A^S = 1-q$, and, by correct conjectures, $\mathbb{E}[D_A, \beta_B | I] = 2(1-q) < \frac{2}{\theta^L}$ given $\theta^L < \bar{\theta}(q)$.
(ii) (I,S,S) with $\alpha_A^I = 1$, $\alpha_A^S = 1$, and $\mathbb{E}[D_A, \beta_B | I] = 2 > \frac{2}{\theta^L}$, given $\theta^L > 1$.

B.1.7. Region VII: $\bar{\theta}(q) \leq \theta^L < 2$ and $\bar{\theta}(q) \leq \theta^H < 2$

Both the low and the high guilt B may find it optimal to Share or Take, depending on A's beliefs. As in Regions IV and V, they can both Take, both Share, or one Share and one Take. A's best response is to go In if at least the high guilt B Shares, and to stay Out if at most the low guilt B does it. Two of these strategy profiles are equilibria:

(i) (O,T,T) with $\alpha_A^I = 0$, $\alpha_A^S = 0$, and, by correct conjectures, $\mathbb{E}[D_A, \beta_B | I] = 1 < \frac{2}{\theta}$ given $\theta^L < 2$ and $\theta^H < 2$.
(ii) (I,S,S) with $\alpha_A^I = 1$, $\alpha_A^S = 1$ and by correct conjectures, $\mathbb{E}[D_A, \beta_B | I] = 2 > \frac{2}{\theta}$ given θ^L and $\theta^H > \frac{1}{1-q}$.

B.1.8. Region VIII: $\theta^L \geq \bar{\theta}(q)$ and $\theta^H \geq 2$

B's best response is to Share if he has high guilt sensitivity. A has the opportunity of securing at least 1 by choosing Out, so she must expect to gain at least 1. Hence, the minimum level of guilt when B Takes after In is $\theta * 1$ which is enough to induce him to Share in this parametric region. As a consequence, also the low guilt player B finds it optimal to Share, as the minimum

disappointment is $2(1-q) > \frac{2}{\theta^L}$. Hence, player A goes In in equilibrium. The pure equilibrium strategies are (I,S,S), and the second-order beliefs are the degenerate ones that obtain from the equilibrium strategies.

B.2. Case II: $q > \frac{1}{2}$

The right panel of Fig. 11 highlights the regions that we consider. Note that when $q < \frac{1}{2}$, $\bar{\theta}(q) > 2$, so that we have a smaller number of regions in this case.

B.2.1. Region I: $\theta^L < 1$ and $\theta^H < 2$

Behavior in this region is as in $q > \frac{1}{2}$, Region I. The pure equilibrium strategies are (O,T,T), and the second-order beliefs are the associated degenerate ones.

B.2.2. Region II: $\theta^L < 1$ and $\theta^H > 2$

B's best response is to Take if he has low guilt sensitivity θ^L, and to Share if he has high guilt sensitivity, independently of his second-order belief. Given $q > \frac{1}{2}$ player A finds it optimal to go Out. Therefore the pure equilibrium strategies are (O,T,S), and the second-order beliefs are the degenerate ones that obtain from the equilibrium strategies.

B.2.3. Region III: $1 < \theta^L < 2$ and $1 < \theta^H < 2$

Both the low and the high guilt B may find it optimal to Share or Take, depending on player A's beliefs. As in $q < \frac{1}{2}$, Region IV, they can both Share, both Take, or one Take and one Share. A's best response is to go In if at least the low guilt B Shares, and to stay Out if at most the high guilt B does it. Two of these strategy profiles are equilibria:

(i) (O,T,T) with $\alpha_A^I = 0$, $\alpha_A^S = 0$, and, by correct conjectures, $\mathbb{E}[D_A, \beta_B | I] = 1 < \frac{2}{\theta}$ given $\theta^L < 2$ and $\theta^H < 2$.
(ii) (I,S,S) with $\alpha_A^I = 1$, $\alpha_A^S = 1$ and by correct conjectures, $\mathbb{E}[D_A, \beta_B | I] = 2 > \frac{2}{\theta}$ given $\theta^L > 1$ and $\theta^H > 1$.

B.2.4. Region IV: $1 < \theta^L < 2$ and $\theta^H > 2$

The high guilt B Shares, while the low guilt may find it optimal to Share or Take, depending on A's beliefs. A's best response is to go In if both high and low guilt players Share, and to stay Out if only the high guilt B does it. The equilibrium strategy profiles are:

(i) (O,T,S) with $\alpha_A^I = 0$, $\alpha_A^S = 1-q$, and, by correct conjectures, $\mathbb{E}[D_A, \beta_B | I] = 1 > \frac{2}{\theta^H}$ given $\theta^H > 2$.
(ii) (I,S,S) with $\alpha_A^I = 1$, $\alpha_A^S = 1$ and by correct conjectures, $\mathbb{E}[D_A, \beta_B | I] = 2 > \frac{2}{\theta}$ given $\theta^L > 1$ and $\theta^H > 2$.

B.2.5. Region V: $\theta^L > 2$ and $\theta^H > 2$

The best response of both high and low guilt B players is to Share. Hence, A goes In. The pure equilibrium strategies are (I,S,S), with the associated second-order beliefs.

B.3. Equilibrium characterization with homophily

When we introduce homophily, also the group of player A becomes relevant. Therefore the profile of equilibrium strategies has now length four.

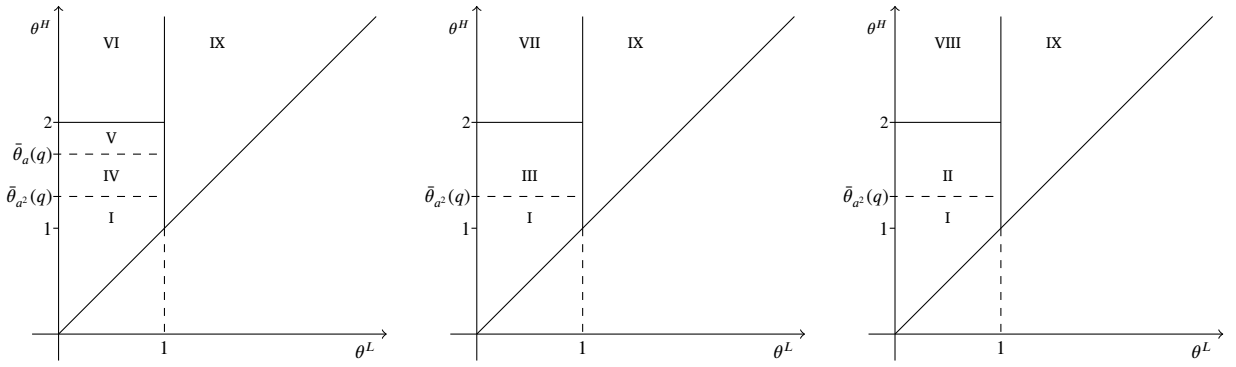
Let us now gather/derive the elements that we need to compute the equilibria Fig. 12 provides a graphical representation of the relevant regions.

1. *Matching probabilities.* The probability that a player from group L is matched with a player from the same group at time t is $\rho_t^L = a + (1-a)q_t$. The probability that a player from group H is matched with a player from the same group at time t is $\rho_t^H = a + (1-a)(1-q_t)$.
2. *A's expectation on matched B.* Player A's probability of being matched with a player B from group L now depends on the group of player A. We denote these probabilities with $\hat{q}_t^k$, where $k = H, L$ denotes the group player A belongs to. These probabilities are:

$$\hat{q}_t^L = \rho^L = a + (1-a)q_t$$
$$\hat{q}_t^H = 1 - \rho^H = (1-a)q_t.$$

3. *B's belief on A's belief on being matched with a B player from group L.* Let us call $\mathbb{E}^k(\hat{q}_t)$ the correct belief of a player B of group k on the expectation of his matched A's on his own group. This belief depends on B's group, as his group affects the probability of being matched with a player A from a specific group, together with the fact that player A's expectations (correctly) depend on her group. Let us compute these two beliefs.

$$\mathbb{E}^H(q_t) = \rho^H \hat{q}_t^H + (1-\rho^H)\hat{q}_t^L = (1-a^2)q_t,$$



(a) $q \leq \frac{1-2a}{2-2a}$ (b) $\frac{1-2a}{2-2a} < q \leq \frac{1}{2-2a}$ (a) $q > \frac{1}{2-2a}$

Fig. 12. Role-dependent guilt, unknown matching, homophily, regions of analysis.

$$\mathbb{E}^L(q_t) = \rho^L \hat{q}_t^L + (1-\rho^L)\hat{q}_t^H = a^2 + (1-a^2)q_t.$$

B.3.1. Region I: $\theta^L < 1$, $\theta^H < \frac{1}{1-(1-a)q}$

B from group L Takes, regardless of his beliefs, as $\theta^L < 1$. A from group i goes In if $\hat{q}^i \alpha_A^{L,S} + (1-\hat{q}^i)\alpha_A^{H,S} \geq \frac{1}{2}$. In this region this is equivalent to $(1-\hat{q}^i)\alpha_A^{H,S} \geq \frac{1}{2}$, as $\alpha_A^{L,S} = 0$. A's disappointment therefore is

$$D_A^i(I,T,\alpha_A) = 1 - \alpha_A^{i,I} + 2(1-\hat{q}^i)\alpha_A^{H,S}\alpha_A^{i,I},$$

which is increasing in α_A^I, $\alpha_A^{H,S}$ and decreasing in $\hat{q}^i$. Note that, in this region $D_A^H(I,T,\alpha_A) > D_A^L(I,T,\alpha_A)$, as A players from group H correctly expect to be matched with a B player from group H more often, and only B players from group H may choose Share with positive probability. B's expectation of player A's disappointment, when player B belongs to group i, is

$$\mathbb{E}^i[D_A] = \hat{q}^i D_A^L + (1-\hat{q}^i)D_A^H,$$

when A players of both groups choose In, $D_A^H(I,T,\alpha_A)$ when only A players from group H choose In, and $D_A^L(I,T,\alpha_A)$ when only A players from group L choose In. The expected disappointment is maximum for B players of group H when: (i) only A players from group H choose In, that is when $\alpha_A^{H,I} = 1$ and $\alpha_A^{L,I} = 0$; and (ii) player A believes that all B players of group H Share, that is $\alpha_A^{H,S} = 1$. In this case, the expected disappointment for a B player of group H is

$$\mathbb{E}^H[D_A(I,T,\alpha_A)] = 2(1-(1-a)q).$$

Even in this case a B player from group H Takes, as $\theta^H < \bar{\theta}_a(q) := \frac{1}{1-(1-a)q}$. Given that every B Takes, every A stays Out, and the equilibrium behavior is (O,O,T,T).

B.3.2. Region II: $\theta^L < 1$, $\frac{1}{1-(1-a)q} < \theta^H < 2$, $q > \frac{1}{2(1-a)}$

B players from group L Take. Given the analysis of Region I, B players of group H Share if only A players of group H go In. However, A players from group H stay out even if all B players of group H Share, because the probability of being matched with a B player of group H is $\rho^H = 1-(1-a)q < \frac{1}{2}$. This holds a fortiori for A players from group L. Therefore the only equilibrium strategy profile in this region is (O,O,T,T).

B.3.3. Region III: $\theta^L < 1$, $\frac{1}{1-(1-a)q} < \theta^H < 2$, $\frac{1-2a}{2-2a} < q < \frac{1}{2(1-a)}$

B players from group L Take. Given the analysis of Region I, B players of group H Share if only A players of group H go In. A players from group H go In if all B players of group H Share, because the probability of being matched with a B player of group H is $\rho^H = 1-(1-a)q > \frac{1}{2}$. A players of group L, instead, stay Out, because their probability of being matched to a B player from group H is $(1-a-(1-a)q) < \frac{1}{2}$. Also (O,O,T,T) is sustainable as equilibrium strategy profile in this region, however it is Pareto-dominated by (O,I,T,S). This can be shown by noting that: (i) a player A from group L has the same utility under both strategy profiles; (ii) a player A from group H has a higher utility under (O,I,T,S), as she can ensure herself the same utility as in (O,O,T,T) by going Out and she chooses not to do so; (iii) a player B from group L has a higher expected utility under (O,I,T,S), as if he is matched to a player A from group H he Takes and experiences a utility higher than 2 (otherwise he would have Shared) instead of the utility of 1 that he receives from player A going Out, which is the only possible outcome under (O,O,T,T); (iv) a player B from group H has a higher expected utility under (O,I,T,S), as if he is matched to a player A from group H he Shares and experiences a utility equal to 2 instead of the utility of 1 that he receives from player A going Out, which is the only possible outcome under (O,O,T,T).

B.3.4. Region IV: $\theta^L < 1$, $\frac{1}{1-(1-a)q} < \theta^H < \frac{1}{1-(1-a^2)q}$, $q < \frac{1-2a}{2-2a}$

B players from group L Take. Given the analysis of Region I, B players of group H Share if only A players of group H go In. A players from group H go In if all B players of group H Share, because the probability of being matched with a B player of group H is $\rho^H = 1-(1-a)q > \frac{1}{2}$. A players of group L go In because their probability of being matched to a B player from group H is $1-\rho^L = (1-a-(1-a)q) > \frac{1}{2}$. However, B players of group H find it optimal to Share if only A players of group H go In, but they do not find it optimal to go In if A players of group L go In as well, as they have lower expectations. In particular, the expected disappointment for a B player of group H if A players of both groups go In is

$$\mathbb{E}^H[D_A(I,T,\alpha_A)] = \hat{q}^H D_A^L(I,T,\alpha_A) + (1-\hat{q}^H) D_A^H((I,T,\alpha_A) = 2(1-(1-a^2)q).$$

Hence, B players from group H do not find it optimal to Share when A-players of both groups go In as long as $\theta^H < \bar{\theta}_{a^2}(q) := \frac{1}{1-(1-a^2)q}$. The only profile of strategies that is sustainable in equilibrium is therefore (O,O,T,T).

B.3.5. Region V: $\theta^L < 1$, $\frac{1}{1-(1-a^2)q} < \theta^H < 2$, $q < \frac{1-2a}{2-2a}$

B players from group L Take. Given the analysis of Region I, B players of group H Share if only A players of group H go In. A players from group H go In if all B players of group H Share, because the probability of being matched with a B player of group H is $\rho^H = 1-(1-a)q > \frac{1}{2}$. A players of group L go In because their probability of being matched to a B player from group H is $1-\rho^L = (1-a-(1-a)q) > \frac{1}{2}$. B players of group H Share even if A players of both groups go In. The expected disappointment for a B player of group H if A players of both groups go In is

$$\mathbb{E}^H[D_A(I,T,\alpha_A)] = \hat{q}^H D_A^L(I,T,\alpha_A) + (1-\hat{q}^H) D_A^H(I,T,\alpha_A) = 2(1-(1-a^2)q).$$

Hence, B players from group H find it optimal to Share when B players of group L Take and A players of both groups go In given that $\theta > \bar{\theta}_{a^2}(q) = \frac{1}{1-(1-a^2)q}$. So, (I,I,T,S) is sustainable as an equilibrium. Note that also (O,O,T,T) is sustainable as equilibrium strategy profile in this region, however it is Pareto-dominated by (I,I,T,S).

B.3.6. Region VI: $\theta^L < 1$, $\theta^H > 2$, $q < \frac{1-2a}{2-2a}$

B players from group L Take and B players of group H Share. A players from group H go In if all B players of group H Share, because the probability of being matched with a B player of group H is $1-(1-a)q > \frac{1}{2}$. A players of group L go In because their probability of being matched to a B player from group H is $(1-a-(1-a)q) > \frac{1}{2}$. Hence, the only equilibrium strategy profile sustainable as an equilibrium is (I,I,T,S).

B.3.7. Region VII: $\theta^L < 1$, $\theta^H > 2$, $\frac{1-2a}{2-2a} < q < \frac{1}{2-2a}$

B players from group L Take and B players of group H Share. A players from group H go In if all B players of group H Share, because the probability of being matched with a B player of group H is $1-(1-a)q > \frac{1}{2}$. A players of group L stay Out because their probability of being matched to a B player from group H is $(1-a-(1-a)q) < \frac{1}{2}$. Hence, the only equilibrium strategy profile can be sustained in equilibrium is (O,I,T,S).

B.3.8. Region VIII: $\theta^L < 1$, $\theta^H > 2$, $q > \frac{1}{2-2a}$

B players from group L Take and B players of group H Share. A players from group H go Out if all B players of group H Share, because the probability of being matched with a B player of group H is $1-(1-a)q < \frac{1}{2}$. A players of group L stay Out because their probability of being matched to a B player from group H is $(1-a-(1-a)q) < \frac{1}{2}$. Hence, the only equilibrium strategy profile which can be sustained in equilibrium is (O,O,T,S).

B.3.9. Region IX: $\theta^H \geq \theta^L > 1$

In this region (I,I,S,S) is an sequential equilibrium, given that in both groups B players find it optimal to Share whenever every A player expects them to do so. If B players of both groups find it optimal to Share, A players from both groups find it optimal to go In. In this region, (I,I,S,S) is also the only equilibrium strategy profile that survives our Pareto-dominance criterion of equilibrium selection, see working paper version of this paper for further details.

References

Alesina, A., Giuliano, P., 2015. Culture and institutions. J. Econ. Lit. 53 (4), 898–944.

Algan, Y., Cahuc, P., 2010. Inherited trust and growth. Am. Econ. Rev. 100 (5), 2060–2092.

Attanasi, G., Battigalli, P., Manzoni, E., 2016. Incomplete-information models of guilt aversion in the trust game. Manag. Sci. 62 (3), 648–667.

Attanasi, G., Battigalli, P., Manzoni, E., Nagel, R., 2019. Belief-dependent preferences and reputation: experimental analysis of a repeated trust game. J. Econ. Behav. Organ. 167, 341–360.

Attanasi, G., Battigalli, P., Manzoni, E., Nagel, R., 2022. Disclosure of belief-dependent preferences in the trust game. IGIER Working paper, 506.

Bandura, A., Walters, R.H., 1963. Social Learning and Personality Development.

Battigalli, P., Comportamento razionale ed equilibrio nei giochi e nelle situazioni sociali.azionale ed equilibrio nei giochi e nelle situazioni sociali. Unpublished undergraduate dissertation, Bocconi University, Milano, 1987.

Battigalli, P., Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M., 2015. Self-confirming equilibrium and model uncertainty. Am. Econ. Rev. 105 (2), 646–677.

Battigalli, P., Corrao, R., Dufwenberg, M., 2019a. Incorporating belief-dependent motivation in games. J. Econ. Behav. Organ.
Battigalli, P., Dufwenberg, M., 2007. Guilt in games. Am. Econ. Rev. Pap. Proc. 97 (2), 170–176.
Battigalli, P., Dufwenberg, M., 2009. Dynamic psychological games. J. Econ. Theory 144 (1), 1–35.
Battigalli, P., Dufwenberg, M., 2022. Belief-dependent motivations and psychological game theory. J. Econ. Lit.
Battigalli, P., Dufwenberg, M., Smith, A., 2019b. Frustration, aggression, and anger in leader-follower games. Games Econ. Behav. 117, 15–39.
Bellemare, C., Sebald, A., Suetens, S., 2017. A note on testing guilt aversion. Games Econ. Behav. 102, 233–239.
Bellemare, C., Sebald, A., Suetens, S., 2019. Guilt aversion in economics and psychology. J. Econ. Psychol. 73, 52–59.
Bentzen, J.S., 2019. Why are some societies more religious than others? In: Advances in the Economics of Religion, pp. 265–281.
Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity, and social history. Games Econ. Behav. 10 (1), 122–142.
Bisin, A., Carvalho, J.-P., Verdier, T., 2023. Cultural transmission and religion. In: The Economics of Religion. World Scientific, pp. 1–62.
Bisin, A., Topa, G., Verdier, T., 2004. Cooperation as a transmitted cultural trait. Ration. Soc. 16 (4), 477–507.
Bisin, A., Verdier, T., 2000. "Beyond the melting pot": cultural transmission, marriage, and the evolution of ethnic and religious traits. Q. J. Econ. 115 (3), 955–988.
Bisin, A., Verdier, T., 2001. The economics of cultural transmission and the dynamics of preferences. J. Econ. Theory 97 (2), 298–319.
Bisin, A., Verdier, T., 2011. The economics of cultural transmission and socialization. In: Benhabib, Jess, Bisin, Alberto, Jackson, Matt (Eds.), Handbook of Social Economics, vol. 1. Elsevier. Chapter 9.
Carvalho, J.-P., 2013. Veiling. Q. J. Econ. 128 (1), 337–370.
Carvalho, J.-P., Iyer, S., Rubin, J., 2019. Advances in the Economics of Religion. Springer.
Chalfant, H.P., Heller, P.L., 1991. Rural/urban versus regional differences in religiosity. Rev. Relig. Res., 76–86.
Charness, G., Dufwenberg, M., 2006. Promises and partnership. Econometrica 74 (6), 1579–1601.
Cheadle, J.E., Schwadel, P., 2012. The 'friendship dynamics of religion,'or the 'religious dynamics of friendship'? A social network analysis of adolescents who attend small schools. Soc. Sci. Res. 41 (5), 1198–1212.
Coleman, J., 1958. Relational analysis: the study of social organizations with survey methods. Human Organ. 17 (4), 28–36.
Currarini, S., Jackson, M.O., Pin, P., 2009. An economic model of friendship: homophily, minorities, and segregation. Econometrica 77 (4), 1003–1045.
Dekel, E., Ely, J.C., Yilankaya, O., 2007. Evolution of preferences. Rev. Econ. Stud. 74 (3), 685–704.
Della Lena, S., Dindo, P., 2023. An economic model of acculturation under strategic complements and substitutes. Unpublished manuscript.
Della Lena, S., Panebianco, F., 2021. Cultural transmission with incomplete information. J. Econ. Theory 198, 105373.
Dohmen, T., Falk, A., Huffman, D., Sunde, U., 2012. The intergenerational transmission of risk and trust attitudes. Rev. Econ. Stud. 79 (2), 645–677.
Durkheim, E., Swain, J.W., 1915. The Elementary Forms of the Religious Life Trans from the French. George Allen and Unwin Limited.
Erikson, E.H., 1993. Childhood and Society. WW Norton & Company.
Erikson, E.H., 1994. Identity and the Life Cycle. WW Norton & Company.
Ferguson, T.J., Stegge, H., Damhuis, I., 1991. Children's understanding of guild and shame. Child Dev. 62 (4), 827–839.
Fudenberg, D., Levine, D.K., 1993. Self-confirming equilibrium. Econometrica, 523–545.
Geanakoplos, J., Pearce, D., Stacchetti, E., 1989. Psychological games and sequential rationality. Games Econ. Behav. 1 (1), 60–79.
Guiso, L., Sapienza, P., Zingales, L., 2003. People's opium? Religion and economic attitudes. J. Monet. Econ. 50 (1), 225–282.
Guiso, L., Sapienza, P., Zingales, L., 2004. The role of social capital in financial development. Am. Econ. Rev. 94 (3), 526–556.
Guiso, L., Sapienza, P., Zingales, L., 2006. Does culture affect economic outcomes? J. Econ. Perspect. 20 (2), 23–48.
Guiso, L., Sapienza, P., Zingales, L., 2008a. Social capital as good culture. J. Eur. Econ. Assoc. 6 (2–3), 295–320.
Guiso, L., Sapienza, P., Zingales, L., 2008b. Trusting the stock market. J. Finance 63 (6), 2557–2600.
Güth, W., Kliemt, H., 1994. Competition or co-operation: on the evolutionary economics of trust, exploitation and moral attitudes. Metroeconomica 45 (2), 155–187.
Güth, W., Yaari, M., 1992. Explaining reciprocal behavior in simple strategic games: an evolutionary approach. In: Explaining Forces and Changes: Approaches to Evolutionary Economics.
Haselton, M.G., Ketelaar, T., 2006. Irrational emotions or emotional wisdom? The evolutionary psychology of affect and social behavior. In: Affect in Social Thinking and Behavior, vol. 8, p. 21.
Iannaccone, L.R., 1998. Introduction to the economics of religion. J. Econ. Lit. 36 (3), 1465–1495.
Iyer, S., 2016. The new economics of religion. J. Econ. Lit. 54 (2), 395–441.
Knack, S., Keefer, P., 1997. Does social capital have an economic payoff? A cross-country investigation. Q. J. Econ. 112 (4), 1251–1288.
Kochanska, G., Gross, J.N., Lin, M.-H., Nichols, K.E., 2002. Guilt in young children: development, determinants, and relations with a broader system of standards. Child Dev. 73 (2), 461–482.
Kreps, D.M., 1990. Corporate culture and economic theory. In: Alt, J.E., Shepsle, K.A. (Eds.), Perspectives on Positive Political Economy. Cambridge Univ. Press, Cambridge, UK, pp. 90–143.
Kreps, D.M., Wilson, R., 1982. Sequential equilibria. Econometrica, 863–894.
La Porta, R., Lopez-de Silanes, F., Shleifer, A., Vishny, R., 1997. Trust in large organizations. Am. Econ. Rev. 87 (2), 333–338.
Levy, G., Razin, R., 2012. Religious beliefs, religious participation, and cooperation. Am. Econ. J. Microecon. 4 (3), 121–151.
Levy, G., Razin, R., 2014. Rituals or good works: social signaling in religious organizations. J. Eur. Econ. Assoc. 12 (5), 1317–1360.
Louch, H., 2000. Personal network integration: transitivity and homophily in strong-tie relations. Soc. Netw. 22 (1), 45–64.
Mannahan, R., 2019. Self-esteem and rational self-handicapping. Unpublished.
Norenzayan, A., 2013. Big Gods. Princeton University Press.
Norenzayan, A., Shariff, A.F., 2008. The origin and evolution of religious prosociality. Science 322 (5898), 58–62.
Norenzayan, A., Shariff, A.F., Gervais, W.M., Willard, A.K., McNamara, R.A., Slingerland, E., Henrich, J., 2016. The cultural evolution of prosocial religions. Behav. Brain Sci. 39, e1.
Okada, A., 2020. The cultural transmission of trust and trustworthiness. J. Econ. Behav. Organ. 169, 53–69.
Oviedo, L., 2016. Religious attitudes and prosocial behavior: a systematic review of published research. Relig. Brain Behav. 6, 169–184.
Patacchini, E., Zenou, Y., 2016. Social networks and parental behavior in the intergenerational transmission of religion. Quant. Econ. 7 (3), 969–995.
Prummer, A., Siedlarek, J.-P., 2017. Community leaders and the preservation of cultural traits. J. Econ. Theory 168, 143–176.
Razin, R., 2019. Religion and segregation. In: Advances in the Economics of Religion, pp. 89–101.
Shariff, A.F., Norenzayan, A., 2007. God is watching you: priming god concepts increases prosocial behavior in an anonymous economic game. Psychol. Sci. 18 (9), 803–809.
Sheldon, K.M., 2006. Research: catholic guilt? Comparing catholics' and protestants' religious motivations. Int. J. Psychol. Relig. 16 (3), 209–223.
Tabellini, G., 2008a. Institutions and culture. J. Eur. Econ. Assoc. 6 (2–3), 255–294.
Tabellini, G., 2008b. The scope of cooperation: values and incentives. Q. J. Econ. 123 (3), 905–950.
Tabellini, G., 2010. Culture and institutions: economic development in the regions of Europe. J. Eur. Econ. Assoc. 8 (4), 677–716.
Tan, J.H., Vogel, C., 2008. Religion and trust: an experimental study. J. Econ. Psychol. 29 (6), 832–848.
Verdier, T., Zenou, Y., 2015. The role of cultural leaders in the transmission of preferences. Econ. Lett. 136, 158–161.
Verdier, T., Zenou, Y., 2018. Cultural leader and the dynamics of assimilation. J. Econ. Theory 175, 374–414.
Walinga, P., Corveleyn, J., van Saane, J., 2005. Guilt and religion: the influence of orthodox protestant and orthodox catholic conceptions of guilt on guilt-experience. Arch. Psychol. Relig. 27 (1), 113–135.
Weber, M., 2013. Religious rejections of the world and their directions. In: From Max Weber. Routledge, pp. 351–387.