ECONOMICS

Working Paper Series

No. 17-2020

Unhappy is the land without symbols – Group symbols in infinitely repeated public good games

Tom Potoms
Department of Economics, University of Sussex Business School
T.Potoms@sussex.ac.uk

Tom Truyts
CEREC, Saint-Louis University
tom.truyts@usaintlouis.be

Abstract: How are group symbols (e.g. a flag, Muslim veil, clothing style) helpful in sustaining cooperation and social norms? We study the role of symbols in an infinitely repeated public goods game with random matching, endogenous partnership termination, limited information flows and endogenous symbol choice. We characterize an efficiently segregating equilibrium, in which players only cooperate with others bearing the same symbol. Players bearing a scarcer symbol face a longer expected search time to find a cooperative partner upon partnership termination, and can therefore sustain higher levels of cooperation. We compare this equilibrium to other equilibria in terms of Pareto dominance and robustness to (some form of) bilateral renegotiation.

Unhappy is the land without symbols - Group symbols in infinitely repeated public good games *

Tom Potoms† and Tom Truyts ‡

September 21, 2020

Abstract

How are group symbols (e.g. a flag, Muslim veil, clothing style) helpful in sustaining cooperation and social norms? We study the role of symbols in an infinitely repeated public goods game with random matching, endogenous partnership termination, limited information flows and endogenous symbol choice. We characterize an efficiently segregating equilbrium, in which players only cooperate with others bearing the same symbol. Players bearing a scarcer symbol face a longer expected search time to find a cooperative partner upon partnership termination, and can therefore sustain higher levels of cooperation. We compare this equilibrium to other equilibria in terms of Pareto dominance and robustness to (some form of) bilateral renegotiation.

Keywords: Endogenous segregation; repeated games; random matching ; public goods games

JEL codes: C73, D83.

1 Introduction

Migration and globalization have given rise to a surge of nationalism, xenophobia and radicalism all over the globe. As such, cultural diversity and societal polarization are high on

*This is a heavily revised paper from an earlier draft circulated under the title "On Symbols and Cooperation". We are grateful to Bram De Rock, Pawel Dziewulski, Matthew Embrey, Paola Manzini, Heiner Schumacher, Ilona Sergant and seminar and conference participants in Brussels, Louvain-la-Neuve, Stony Brook, Paris, Oxford, Utrecht, Leuven, Saint Petersburg and Maastricht. Tom Truyts gratefully acknowledges financial support from the Belgian French speaking community ARC project n15/20-072, "Social and Economic Network Formation under Limited Farsightedness: Theory and Applications", Universités Saint-Louis - Bruxelles, October 2015 - September 2020.

†University of Sussex. Jubilee building, Falmer Brighton BN1 9SL, United Kingdom. E-mail: T.Potoms@sussex.ac.uk

‡CEREC, Saint-Louis University - Brussels, Boulevard du Jardin botanique 43, 1000 Brussels, Belgium; Center for Economic Studies, University of Leuven. E-mail: tom.truyts@usaintlouis.be, tel.: +3222117845.

the policy agenda in most countries. The social tensions and debates often focus on various cultural markers or group symbols, as illustrated by the ban of the Muslim veil in several European countries, or the recent polarization surrounding the wearing of face coverings during the COVID pandemic[1]. But the exact importance and functioning of these group symbols remains more obscure. These group symbols, e.g., a flag, a Muslim veil, a T-shirt of a rock band or an expensive corporate style suit, function in various ways as coordination devices. On one hand, they reveal information about underlying heterogeneity. Strangers can form a reasonably accurate idea about one's socioeconomic background and tastes from a casual observation of one's clothing and lifestyle. [2]

On the other hand, symbols strengthen group identification and loyalty. By displaying the above symbols, one is met with initial sympathy from some strangers, and with aversion from others. Tajfel and Turner (1979) their famous 'minimal group experiment' shows that symbols can give rise to a differential sympathy or hostility towards strangers, even if these symbols are understood to reflect no underlying heterogeneity. [3] While these findings gave rise to an extensive body of literature and numerous replications, the underlying mechanisms are relatively poorly understood.

Iannaccone (1992) presents an interesting interpretation of the role of symbols in the context of cults and sects. He understands these symbols as a solution for a typical group problem: the underprovision of a club good. The standard solution in club theory is to levy membership fees, and then use the revenues to subsidize contributions to the club good (see e.g., Sandler and Tschirhart (1997)). But if such a formal scheme is unfeasible or undesirable, Iannaccone (1992) suggests, then a cult or sect can 'tax' resources spent outside the group. If members contribute time to the club good, such a 'tax' takes the form of restrictions on members' clothing, diet, haircut or language, all of which impede social interactions with non-members. By sacrificing their capacity for social interactions outside the group, these

[1]There are many reports in the news that indicate support for face coverings in public spaces breaks down very much along partisan lines. For example, according to a NBC News/Survey Monkey Weekly Tracking Poll (July 20-26, 2020) almost 97 percent of Democrats or those who lean Democratic indicate they wear masks at least most of the time when they leave the house, whereas this drops to approximately 70 percent of Republicans (with only 48 percent of Republican-registered respondents indicating they wear masks 'all the time' when they leave the home, in comparison to 86 percent for Democrats and 71 percent of Independents.)

[2]Even seemingly innocuous symbols can operate as a signal of 'trustworthiness' or 'friendliness' and can evoke feelings of sympathy and thereby consequent strategic interactions, e.g., Manzini et al. (2009) show that, exchanging a smiley or wink emoji to indicate that each player is ready to play a minimum effort game affects their equilibrium efforts.

[3]Tajfel and Turner (1979) randomly allocated a minimal marker to test subjects, e.g. a colored dot on the forehead, and found that such a minimal symbol sufficed to cause participants to discriminate between complete strangers on the basis of their symbol, even in the absence of any underlying heterogeneity. Test subjects bestowed advantages on strangers bearing the same symbol, at the expense of others bearing a different symbol.

typical idiosyncrasies of religious, political or subcultural groups help members to commit their resources to the group.

The sacrifice of outside options in order to demonstrate commitment and sustain cooperation and group norms is documented by social scientists in a variety of contexts. Gambetta (2011) discusses how prisoners demonstrate their commitment to a life in crime by applying prison tattoos on visible body parts, thus ruining their chances of an honest life. Gambetta (2011) equally describes how candidate members of Colombian youth gangs are required to kill a friend or family member. Besides proving one's ability to murder, it also shatters gang members' fall back option for leaving the gang. Berman (2000) documents these sacrifice mechanisms for the case of ultra-orthodox Jews. Berndt (2007) shows how being a member of a distinct and despised ethnic or religious minority, and the implied lack of outside options, allowed e.g., 19th century Jewish peddlers to act as middlemen in high stake financial transactions. Shimizu (2011) models self-sacrifice in military and terrorist groups as a result of giving up individual autonomy. Aimone et al. (2013) find that the possibility of sacrificing private outside options enhances club good contributions in a Voluntary Contribution Mechanism experiment.

What we do. In this paper, we study the role of symbols in the context of an infinitely repeated public goods game, with random matching, endogenous partnership termination and limited information flows. We focus on stationary public perfect equilibria (PPE) of the game. We consider an infinite population of homogeneous players, who differ *ex ante* only in a visible but payoff-irrelevant symbol (e.g., a colored hat). Players begin each round with one partner, with whom they play a stage game consisting of two phases. First, they play a public goods game (with continuous effort choices). Second, upon observing the public goods game's outcome, both players simultaneously decide whether to terminate the partnership or not, and whether to change their symbol at a certain cost. Partnerships break up if at least one partner wishes to terminate, and are otherwise terminated exogenously with a small probability. Furthermore, players whose partnership was terminated are then randomly rematched. Starting a new partnership, players have no information about their partner's past play, but only observe his symbol.[4]

We characterize a class of efficiently segregating equilibrium of this game, in which

- players exert no effort in the public goods game if their partner bears a different symbol.

[4]Notice that random matching, though quite standard, is also very natural in the present setting. Indeed, random matching confronts players with a stream of opportunities to form a beneficial partnership with other players (including those who bear a different symbol) and therefore, allows us to place restrictions on the extent of commitment of players to their partnerships and the segregating equilibrium.

- In partnerships which are homogeneous in terms of symbols, players exert the maximal incentive compatible effort.
- Failure to comply with the equilibrium effort in a homogeneous partnership is punished with partnership termination, thus implying in expectation a certain search time to find a new identical symbol partner to start cooperating with.

Given that in heterogeneous (in terms of symbols) partnerships there are no positive efforts, they constitute a waste of time, and are immediately terminated by both players. Another feature of these sort of equilibria is that players bearing a more scarce symbol face in expectation a longer search for a cooperative partner after a break-up, and this sacrifice of outside options allows them - in the spirit of Iannaccone (1992) - to sustain higher cooperation levels (see our Proposition 1) . This damaging effect on members' outside options (i.e., non-members' reactions to these symbols) is a crucial mechanism for symbols to discipline group members' behavior. A nice feature of our framework is that, in contrast to the literature inspired by Iannaccone (1992), we do not take the negative reaction to group symbols as exogenously given. In particular, Proposition 2 provides the conditions on switching costs and continuation values under which no player would want to switch symbols. This extension towards endogenous symbol choice and reactions as equilibrium behavior is an important extension to real-world applications, since even though a negative reaction is inherent in some cases, such as killing family members, it is much less obvious for more arbitrary and minimal symbols, such as clothing or hair color. To the best of our knowledge, we are the first paper to study cooperative behavior with continuous action spaces, within an environment of limited information flows, repeated random matching and where symbol choices, the reactions to symbols and the resulting cooperation levels, are all jointly derived from (a notion of) equilibrium behavior.

We provide several results comparing the class of efficiently segregating PPE with another class of ('symbol-blind') cooperative PPE which are discussed in the literature. In particular, Proposition 5 shows that, for any symbol-blind PPE we can find a Pareto dominating efficiently segregating PPE. Proposition 6 strengthens this result in the sense that it shows even for the 'best' symbol-blind PPE there exist Pareto dominant efficiently segregating PPE. Furthermore, Propositions 8 and 9 show that the class of efficiently segregating PPE are more robust to joint deviations from equilibrium behavior within partnerships, compared to symbol-blind PPE. These results provide a strong motivation to further analyze the class of efficiently segregating PPE and the role of symbols in environments with repeated random matching and limited information flows.

Finally, we show how the average payoffs (across all symbols) evolve with respect to the distribution of players across symbols. In particular, we show (cfr. Proposition 7) that the uniform distribution does not, in general, yield the largest average payoffs, especially not in societies with a large number of (available) symbols.

Our contribution to the literature. This paper relates to a large body of literature on cooperation in infinitely repeated public goods or prisoner's dilemma games. The central question in this literature is how to constrain the continuation payoffs of defectors in order to sustain cooperation on the equilibrium path, despite of defection being the stage game's dominant strategy. However, the present setting excludes a large number of well-known mechanisms that sustain cooperation. First, endogenous partnership termination and random rematching excludes the entire class of personal enforcement mechanisms, in which cheating triggers a punishment by the victim. Because defectors can terminate a partnership before undergoing their punishment, the usual folk theorems and trigger strategy results do no apply. Second, the absence of information about a partner's past play in previous partnerships excludes community enforcement mechanisms, in which shirkers are identified and punished by other members of the population.[5] Whereas knowledge of the full histories of players might be a plausible assumption in small communities, this is not the case for (relatively) large communities. Since we are interested in the scope for cooperation in large anonymous societies where individuals don't have full access to each other's past interactions, we don't make the assumption that players can observe private histories. Third, even though the contagion mechanisms of Kandori (1992) and Ellison (1994) can sustain cooperation if players are randomly rematched and only aware of their own history of play, they are excluded in this setting by the continuum population.[6]

The literature has nevertheless advanced two mechanisms to sustain cooperation in the present restrictive setting.

1. *Gradual trust-building strategies*, in equilibria which are based on such strategies, partners only engage in full cooperation after a sufficiently long trust-building or 'incubation' phase, i.e., a number of rounds in which they defect or exert low effort. The prospect of a trust-building stage with a new partner suffices to deter players from

[5]Sustaining cooperation through punishments by other community members has been shown effective under various information assumptions by e.g. Greif (1993), Okuno-Fujiwara and Postlewaite (1995), Mailath et al. (2000) or Takahashi (2010). More recent work has focussed on intermediate settings in which players can observe partial information of past behavior of other players, see e.g. Heller and Mohlin (2018), Bhaskar and Thomas (2019), Deb and González-Díaz (2019) and Deb (2020).

[6]In these equilibria, players defect in all their future partnerships if their partner cheats. If players are sufficiently patient, they are dissuaded from defecting by the foresight of eventually triggering the entire population to defect forever. In the present setting, a defection eventually infects at most countably many out of uncountably many players into defecting.

cheating in the later stages of a partnership.[7]

2. *Committed players.* The presence of exogenous defectors in the population gives the situation of having a cooperative partner sufficient scarcity value to discourage defection.[8] Ghosh and Ray (1996) show how cooperation in a public goods game is sustainable if the defectors' population share is neither too small nor too large. Adverse selection, due to the defectors always having to draw a new partner while patient cooperators lock themselves into long term partnerships, means that a small population share of defectors can suffice to sustain cooperation among patient players. More recently, in a richer information framework where players can observe part of their partner's past play, Heller and Mohlin (2018) show that the presence of 'commitment types' can induce cooperative behavior, depending on the strategic environment (type of Prisoner's dilemma).

The present paper also contributes to this literature in the sense that we study a setting similar to Ghosh and Ray (1996)'s repeated public goods game, but in which the role of the exogenous defectors is played in equilibrium by endogenous group symbols. Hence, we assume no preference heterogeneity (in contrast to 'committed players'), but rather derive that players act in equilibrium much like defectors towards others bearing a different symbol. In this equilibrium, players bearing different symbols face generically different incentives. In the spirit of Iannaccone (1992), players can thus sacrifice their outside options by bearing a more scarce symbol, and this sacrifice allows them to sustain higher cooperation levels.

The importance of payoff irrelevant group symbols for cooperation is also central in Eeckhout (2006) and Choy (2018).[9] Eeckhout (2006) studies a public correlation device such as skin color in an infinitely repeated prisoner's dilemma with endogenous partnership termination and limited information. Eeckhout compares a standard ('color-blind') incubation equilibrium to a 'segregation equilibrium', in which new partners of the same color start cooperating immediately, while other new partners play an incubation strategy. Eeckhout shows that color distributions exist for which the segregation equilibrium Pareto dominates the color-blind equilibrium.

Choy (2018) is the closest to our paper. He studies how segregation on the basis of visible group affiliations helps to sustain cooperation in an infinitely repeated public goods game.

[7] See e.g. Datta (1996), Kranton (1996), Eeckhout (2006), Fujiwara-Greve and Okuno-Fujiwara (2009), Fujiwara-Greve et al. (2012). This approach also relates to the idea of 'starting small' in Watson (1999) and Watson (2002), where the stakes of the game gradually increase with the partnership's age.

[8] Related mechanisms are also studied by e.g. Fujiwara-Greve and Okuno-Fujiwara (2009) and Schumacher (2013).

[9] See also Peski and Szentes (2013) on how payoff irrelevant symbols can lead to discriminatory behavior.

Choy assumes that players also know the group affiliation of their partners' previous partners and that groups are hierarchically ranked. He characterizes a renegotiation proof segregating equilibrium, in which players refuse to interact with members of lower groups to protect their reputation. Preserving this reputation implies higher search costs upon partnership termination, which in turn helps to sustain more cooperation. In contrast with Choy (2018), we assume no information about a partner's past play, and unlike Eeckhout (2006) and Choy (2018), we consider symbols a choice variable.

Structure of the paper. The remainder of this paper is organized as follows. The formal setting and equilibrium concept are introduced in Section 2. Section 3 discusses how symbols are helpful in sustaining cooperation by characterizing a class of efficiently segregating equilibria. Section 4 introduces the class of gradual trust-building equilibria in our framework and compares this class with the efficiently segregating PPE. In section 5, we discuss the optimal distribution of symbols in society in terms of average payoffs. Section 6 discusses the robustness of the class of efficiently segregating and the gradual trust-building PPE to (a notion of) bilateral rationality, i.e., joint deviations from equilibrium behavior within partnerships. Finally, section 7 concludes the paper. The main proofs and derivations are detailed in the Appendix.

2 Formal Setting

We assume a continuum of players. Time is discrete and indexed by $t \in \mathbb{N}$, and all players have the same discount factor $0 \leq \delta < 1$. Each player wears one publicly visible symbol out of a given set of symbols $S = \{s^i\}_{i=1,\ldots,n}$. In the initial period, each player is endowed with a particular symbol, after which they can switch their symbol. The fraction of players bearing symbol s^i at the beginning of period t is then denoted by α_t^i. Players start each round of the game with one partner. We call a partnership between two players homogeneous if both bear the same symbol, and we otherwise call it heterogeneous.

The game proceeds as follows: in each round of the game, players first play a public goods game with their current partner. After observing the public goods game's outcome and inferring their partner's contribution, players decide on symbol change and then partnership termination. The symbol switching cost of a player who begins a round t with symbol s^i and ends it with symbol s^j is denoted $c_t(i,j)$, with $c_t(.) \geq 0$ and $c_t(i,j) = 0$ if $i = j$. For simplicity, we assume that these switching costs are independent of t and we collect these switching costs in the matrix $c = (c(i,j))_{i,j}$. Partnerships are exogenously terminated with probability $\lambda \in (0,1]$. Otherwise, both players choose whether or not to continue the partnership, $l \in \{0,1\}$, where $l = 1$ means continuing the partnership. A partnership ends if at least one of the partners wishes to terminate it. If their partnership is terminated, players randomly draw a new partner from the set of players whose partnership was terminated with uniform probability. Of course, the assumption of exogenous partnership termination ensures that drawing a new partner is uninformative about past behavior on the equilibrium path.[10] When meeting a new partner, players do not observe his past behavior but only see his symbol.

We first formally characterize the public goods game. In this game, both partners choose a level of efforts $e \in \mathbb{R}_+$. A player who contributes e while his partner contributes e' obtains a stage payoff $\pi(e, e')$. If we let π_k and π_{kh} denote the partial derivative of π with respect to argument k and with respect to arguments k and h, respectively, then the following restrictions on the public goods game technology are imposed:

Condition 1 *Let $\pi : \mathbb{R}^2 \to \mathbb{R}$ be a function such that:*

1. *π is a twice continuously differentiable function.*

2. *(Public goods game) $\pi_1(e, e') < 0$ for all $e, e' > 0$, $\pi_2(e, e') > 0$ for all $e, e' > 0$, and $\pi_1(e, e) + \pi_2(e, e) > 0$ for all $e \geq 0$,*

3. *(Boundedness) there is a $M \in \mathbb{R}$ s.t. for all $e \in \mathbb{R}_+$, $\pi(e, e) < M$ and $\exists \gamma > 0$ such that $\pi_2(e, e') > \gamma$, for all $e, e' \in \mathbb{R}_+$*

4. *(Initial condition) let $\pi(0, 0) = 0$, while $\pi_1(0, .) = 0$, and $\pi_{11}(0, 0) < 0$.*

The first part of condition 1 ensures smoothness of the payoff function. The second part ensures that π represents a public goods game: the payoffs are decreasing with own effort, such that zero effort is a dominant strategy in the stage game, and increasing with the partner's effort. Moreover, coordinating symmetrically on a higher effort level is always mutually beneficial. The next two parts of condition 1 impose some regularity conditions to ensure that the players' problem and behavior is always well defined. Part 3 of condition 1 bounds the benefits of symmetric efforts, implying that $\lim_{e\to\infty} \pi_1(e, e) + \pi_2(e, e) = 0$, and bounds marginal benefits of the partner's efforts away from zero. Part 4 of condition 1

[10] Note that, in absence of an exogenous break up probability, we would not observe any dissolutions of partnerships on the equilibrium path, hence beliefs within a new formed partnership regarding whether the other deviated from cooperation in his previous partnership are not determined. While in the case of a positive exogenous break up probability, these beliefs are well determined on the equilibrium path. Indeed, the probability that each partner deviated from cooperation in their previous partnership is exactly equal to zero.

normalizes π to be zero in the absence of any contribution and ensures that our problem is well defined near zero. The following simple example shows a public goods game technology which satisfies the above condition and will serve as a closed form example in the remainder of this text.

Example 1 *The payoff function*

$$\pi(e, e') = 1 + e' - e - \frac{1}{1+e}$$

satisfies condition 1, as $\pi_1(e, .) = -1 + \frac{1}{(1+e)^2} < 0$ *for* $e > 0$, $\pi_2(.) = 1 > 0$ *and* $\pi_1(e, e) + \pi_2(e, e) = \frac{1}{(1+e)^2} > 0$ *for all* $e \in \mathbb{R}_+$*. Moreover,* $\pi(e, e) = \frac{e}{1+e}$ *is bounded from above by 1,* $\pi(0, 0) = 0$ *and* $\pi_1(0, .) = 0$.

The information flows between players are limited to public information and everything that has occurred within their present partnership. More formally, a player at the beginning of round t, denoted $h_t \in H_t$, consists of the fundamentals of the game, the symbol of his current partner, and their history of play in the current partnership.[11] Let $H \equiv \bigcup_t H_t$ be the set of all possible information sets. Similar to Eeckhout (2006), we will focus on strategies that are conditional on H and refer to these as public strategies.[12]

A pure public strategy $\sigma : H \to \mathbb{R}_+ \times \{0, 1\}^{\mathbb{R}_+} \times S^{\mathbb{R}_+}$ then specifies a triplet (e_t, l_t, s_t) for all t and possible information sets h_t.[13] The first element specifies how much effort to exert given the symbol of the current partner and the history of play. The second element specifies, for each h_t, a termination decision for all possible effort levels of the partner e'_t:

$$l_t : \mathbb{R}_+ \to \{0, 1\}.$$

[11] Hence, these information assumptions keep players from conditioning their behavior on what happened in their previous partnerships. On one hand, note that players run into one of their past partners again with zero probability, because even for $t \to \infty$, players have met at most countably many out of uncountably many players. On the other hand, our equilibrium concept excludes players from conditioning their behavior on private histories of play (cfr. infra).

[12] Note that allowing players to condition their choices on their private histories of play, i.e., their play in previous partnerships, would introduce asymmetric information into the game. This would complicate the analysis considerably, as it would require us to introduce beliefs about a partner's private information as well as to have strategies equally depending on these beliefs.

[13] As usual, Y^X represents the set of all mappings from X to Y. Note that this formulation is equivalent with players making termination and symbol switching decisions at the second phase of round t as a function of an intermediate history of play, which comprises h_t and the effort strategies in round t's public goods game. A partner's effort choice constitutes the only new information at this intermediate stage of round t.

Similarly, the last element specifies a symbol switching decision as a function of the partner's effort e'_t.[14]

Players evaluate a strategy by considering the expected future payoff streams to which a strategy is expected to give rise; i.e., they wish to maximize

$$\mathbb{E}\left(\sum_t \delta^t \left(\pi\left(e_t, e'_t\right) - c_t\left(i, j\right)\right)\right),$$

in which the expectation operator $\mathbb{E}$ indicates the expectations over all possible future histories of play and symbols of partners to which a strategy σ may lead, given the strategies of other players as well as the stochastic processes of partnership termination and formation. We study the stationary perfect public equilibria (PPE) of this game, i.e., profiles of public strategies which yield for all t and all h_t a Nash equilibrium for round t and all consecutive rounds.

3 The role of symbols for cooperation

This Section discusses how the payoff irrelevant symbols can help to sustain cooperative behavior on the parts of players within (randomly matched) partnerships. To be more precise, we will characterizes a particular class of stationary PPE in pure strategies, which we will refer to as 'efficiently segregating PPE', and which will be the main focus in this paper. We start with a formal description of this class of equilibria:

Definition 1 (Efficient segregating PPE) *The class of efficiently segregating PPE are the stationary PPE in pure strategies in which players*

1. *exert effort 0 in heterogeneous partnerships and, for all i, the maximal effort level which is incentive compatible (i.e. robust to unilateral deviation from each partner within the current partnership), denoted $\bar{e}^i$, in homogeneous s^i partnerships,*

2. *never switch symbol,*

3. *never terminate a homogeneous partnership on the equilibrium path but always terminate a heterogeneous partnership,*

[14] Notice that allowing for players' decisions to terminate their partnerships to depend on their own efforts doesn't make a difference, as own efforts (contingent on their partner's efforts) and the termination decision are made jointly. Therefore any explicit dependency of termination decisions on own efforts is already captured in the joint effort/termination decision.

4. *terminate a partnership after any deviation from equilibrium play, including deviations from equilibrium reactions to deviations, etc.*

An important characteristic of this class of efficiently segregating PPE is that players only exert positive efforts in a homogeneous partnership.[15] As a result, a heterogeneous partnership is a waste of time and is thus terminated immediately. Although players never switch symbols in equilibrium, the ability to switch symbols is relevant, because it allows us to characterize how symbol switching costs determine feasible equilibrium symbol frequencies, and ultimately which patterns of cooperation can be sustained in equilibrium. To simplify a formal discussion of the latter, let $\mathcal{U} = \{\alpha \in [0,1]^n | \sum_{i=1}^n \alpha^i = 1\}$ denote the unit simplex, where we recall that α^i denotes the proportion of players bearing symbol s^i.[16] We will denote with $\mathcal{E}(c) \subseteq \mathcal{U}$ the set of all distributions $\alpha = (\alpha^1, ..., \alpha^n)$ that allow for an efficiently segregating PPE, given the matrix of switching costs. Notice that, when the switching costs become large, then we should expect the set $\mathcal{E}(c)$ to increase, i.e. to encompass more symbol frequencies α that can be made compatible with an efficiently segregating PPE. In contrast, when the switching costs become smaller, then switching becomes easier and we might expect the set of viable symbol frequencies to shrink. We will return to these remarks when we derived formal conditions on frequencies over symbols to be compatible with an efficiently segregating PPE. In the remainder of this section, we proceed in steps to characterize the set of all distributions that are compatible with an efficiently segregating PPE, $\mathcal{E}(c)$.

3.1 Dynamics of symbols and rematching probabilities

Before we can discuss equilibrium effort levels in the efficiently segregating PPE, we need to discuss the (stationary) distribution of symbols in the population and more precisely, we need to assess the probability of rematching with a same symbol player. To that end, we need to introduce a bit more notation. In particular, let x_t^i denote the fraction of players bearing symbol s^i that draw a new partner in period t. The probability, p_t^i that a player with symbol s^i will draw a same-symbol player is then given by

$$p_t^i = \frac{x_t^i}{\sum_j x_t^j}. \tag{1}$$

[15]Of course, equivalent equilibria can be conceived in which players treat a subset of symbols as if it were the same symbol. In this case, one can easily understand the elements of S as partitions of a larger set of symbols.

[16]Notice that we dropped the time subscript as we focus on equilibria in which no player wants to change their original symbols and switch to a different one.

By Definition 1, no player switches symbol in equilibrium (given the switching costs $c(.,.)$), hence, $\alpha_t^i = \alpha^i$ for all t. The dynamics of the shares $(x_t^1, ..., x_t^n)$ are described by the following system of equations:

$$x_{t+1}^i = \left(1 - p_t^i (1 - \lambda)\right) x_t^i + \lambda \left(\alpha^i - x_t^i\right), \text{ for all } i = 1, ..., n. \tag{2}$$

Symbol s^i players currently in a homogeneous partnership only have to draw a new partner if their partnership was exogenously terminated (with probability λ), and a fraction $1-(1-\lambda)p^i$ of s^i players who play the present round with a new partner will have to do so as well in the next round. Since our equilibrium concept requires stationarity, we solve for continuation values from an efficiently segregating PPE by setting $x_t^i = x^i$ for all t, hence $p_t^i = p^i$. We can appeal to Brouwer's fixed point theorem[17] to show that the system of equations represented in (2) has a fixed point, corresponding to a stationary fraction of singles across symbols, x^i, $i = 1, ..., n$, from which we can also derive the stationary (re-)matching probabilities p^i, $i = 1, ..., n$. Now, after some rearranging of (2) we obtain:

$$\left[(1 - \lambda) \frac{x^i}{\sum_j x^j} + \lambda\right] x^i = \lambda \alpha^i. \tag{3}$$

Notice that (3) implies that an increase in α^i which is compensated by decreasing the frequencies of the other symbols, i.e. decreasing $\alpha^j, j \neq i$ increases p^i. However, we note that, players bearing a scarce symbol are are overrepresented in the set of players drawing a new partner (i.e., p^i overstates their overall population share). This is a result of their relative scarcity within the population, thereby increasing the size of such players within the pool of players having to redraw a new partner. Note that this selection effect differs from the adverse selection effect in Ghosh and Ray (1996), where the myopic types are overrepresented in the set of players drawing a new partner, because patient players lock themselves into long-term cooperative partnerships, unlike myopic players. In the efficient segregating PPE, the selection effect softens any asymmetries in symbol frequencies. Because they spend more time looking for a same symbol partner, players bearing a scarce symbol have more chance of finding a same symbol partner in the pool of players looking for a new partner than in the population at large.

[17] In this case, we can apply Brouwer's fixed point theorem to the domain $[\epsilon, 1] \times ... \times [\epsilon, 1] = [\epsilon, 1]^n$ for a sufficiently small $\epsilon > 0$. We do note that (2) does not define a contraction, hence we cannot directly use Banach's fixed point theorem to show unicity of a solution. However, in practice, we conducted numerical simulations and these never presented multiple steady states.

3.2 Effort levels in the efficiently segregating PPE

We now turn our attention to the equilibrium effort levels in the efficiently segregating PPE. The efficiency requirement in Definition 1 implies that, for the particular class of efficiently segregating equilbria players will exert the maximal level of effort which can be sustained in a homogeneous partnership. Note that this form of efficiency is stronger than Pareto efficiency, because a strategy profile in which two partners respectively exert efforts on and strictly below the incentive compatibility constraint is Pareto efficient, but not efficient in the above sense. Since two players bearing the same symbol face the same incentives, the efficient segregating PPE is symmetric in the sense that two players in a homogeneous partnership exert the same effort. This simplifies the analysis, since we only need to compute the expected continuation values of a player (bearing a particular symbol, say s^i) for a given effort level e in a partnership (either heterogeneous or homogeneous). We can use the probabilities of being rematched with a same-symbol bearing player, p^i, $i = 1, ..., n$, i.e., the stationary values from (1) to compute the expected continuation value of an s^i player in a homogeneous partnership recursively as follows:

$$v^i\left(\bar{e}^i\right) = \pi\left(\bar{e}^i, \bar{e}^i\right) + \delta\left(1 - \lambda\right) v^i\left(\bar{e}^i\right) + \delta\lambda w^i\left(\bar{e}^i\right), \tag{4}$$

in which $w^i\left(\bar{e}^i\right)$ is the expected continuation value of an s^i player starting the round with a randomly drawn partner:

$$w^i\left(\bar{e}^i\right) = p^i v^i\left(\bar{e}^i\right) + \left(1 - p^i\right)\delta w^i\left(\bar{e}^i\right). \tag{5}$$

Hence, in equilibrium both players get a stage payoff $\pi\left(\bar{e}^i, \bar{e}^i\right)$ in a homogeneous s^i partnership, after which their partnership survives to the next round with probability $1 - \lambda$, and is terminated otherwise. In case of termination, the s^i players immediately draw a new s^i partner with probability p^i, in which case they start cooperating immediately and thus return to continuation value $v^i\left(\bar{e}^i\right)$. Otherwise they get stage payoff zero, terminate the partnership, and start the next round again with a randomly drawn partner, having continuation value $w^i\left(\bar{e}^i\right)$.

The effort levels which can be sustained as part of a segregating PPE have to satisfy the following incentive compatibility constraints:

$$v^i\left(\bar{e}^i\right) \geq \pi\left(0, \bar{e}^i\right) + \delta w^i\left(\bar{e}^i\right), \tag{6}$$

which states that a s^i player is not worse off when providing the equilibrium effort level

$\bar{e}^i$, rather than defecting on his partner and starting anew with a new partner in the next period. Efficiency then imposes the inequality in (6) to be satisfied with equality. Solving (4) and (5) for v^i and w^i and substituting into (6), we define

$$\begin{aligned} d\left(e,p^i\right) &\equiv v^i\left(e\right)-\pi\left(0,e\right)-\delta w^i\left(e\right) \\ &= \frac{\pi\left(e,e\right)}{1-\delta\left(1-\lambda\right)\left(1-p^i\right)}-\pi\left(0,e\right), \end{aligned} \tag{7}$$

such that (6) can be written as $d\left(\bar{e}^i,p^i\right)\geq 0$. Note that

$$d\left(e,0\right)=\frac{\pi\left(e,e\right)}{1-\delta\left(1-\lambda\right)}-\pi\left(0,e\right) \tag{8}$$

is the difference between the expected actual value of the current partnership, when cooperating at effort level e, and the one shot payoff of cheating. Notice that if players have no hope of finding a new cooperative partner after break-up, e.g., because $p^i=0$, then incentive compatibility requires $d\left(e,0\right)\geq 0$. However, note that d decreases with p^i. Indeed, the possibility of a new partnership with another cooperative s^i player decreases the punishment that breaking up the present partnership constitutes, and it thus reduces the effort levels players can commit to. As a result, we obtain the following effort levels in the efficient segregating PPE.

Proposition 1 *In the efficient segregating PPE, the equilibrium effort in homogeneous s^i partnerships, $\bar{e}^i$, uniquely solves*

$$\bar{e}^i=\max\left\{e|d\left(e,p^i\right)=0\right\}. \tag{9}$$

Moreover, $\bar{e}^i=\bar{e}^i\left(p^i;\lambda,\delta\right)$ is a left-continuous and strictly decreasing function of p^i and λ, and a right-continuous and strictly increasing function of δ.

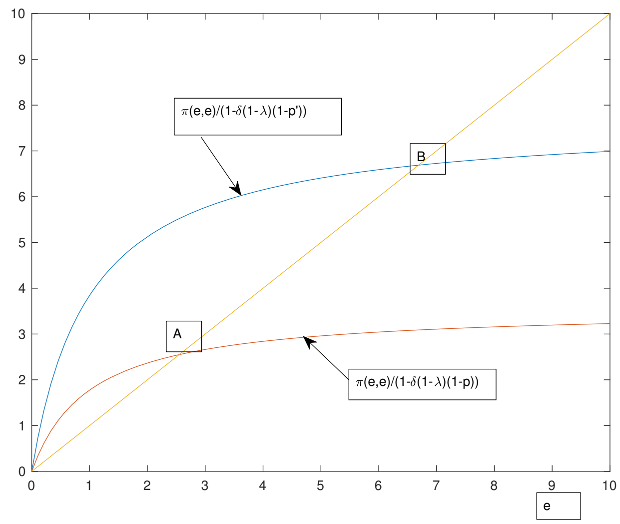
Hence, players with more scarce symbols face a worse outside option, and this allows them to sustain higher effort levels. Moreover, effort levels are increasing with the 'effective' discount factor $\delta\left(1-\lambda\right)$, i.e., if the value of a current homogeneous partnership increases because players become more patient or because the expected longevity of their present partnership increases.

Example 2 *In the closed form example, $\frac{\pi(e,e)}{\pi(0,e)}=\frac{1}{1+e}$ strictly decreases with e, such that $\bar{e}^i$ is a continuous function of p^i. The equilibrium effort levels are*

$$\bar{e}^i=\frac{\left(1-p^i\right)\delta\left(1-\lambda\right)}{1-\left(1-p^i\right)\delta\left(1-\lambda\right)}.$$

We can also illustrate the dependence of efforts on p^i graphically,

Figure 1: Efforts in the efficiently segregating PPE



Efforts for $p^i = p$ will be determined by the intersection point A, whereas (higher) efforts for $p^i = p' < p$ can be read off using intersection point B.

3.3 Symbol switching

We now turn to an analysis of the symbol switching. In particular, given efforts (9) and continuation values (v^i and w^i) we are interested in how the matrix of switching costs determines the set of symbol frequencies $\alpha \in \mathcal{E}(c)$, that can be sustained in an efficiently segregating PPE. Recall that Definition 1 requires that, in any efficient segregating PPE, no player switches symbol. Only in this case is it possible for scarce symbol bearing players to sustain higher efforts (due to longer expected search times). Consequently, the matrix of symbol switching costs, $(c(i,j))_{i,j=1,\ldots,n}$ determines which set of population shares over the different symbols, α are compatible with an efficient segregating PPE. Players can unilaterally switch symbols in the following cases: when in a homogeneous partnership and exerting equilibrium efforts, when in a homogeneous partnership after deviating from the equilibrium effort, and when currently in a heterogeneous partnership. The two latter options give exactly the same continuation values and are therefore payoff equivalent. Furthermore the first option is dominated by these two latter options: a player in a homogeneous s^i partnership who exerts $\bar{e}^i$ and then switches symbol without breaking up will face a partner exerting zero effort and breaking up the partnership in the next round.[18] Therefore the player can strictly

[18]Given that equilibrium efforts are required to be subgame perfect, this implies that a deviating reaction to symbol switching should correspond to a partner exerting zero efforts and terminating the partnership,

improve payoffs by exerting no effort, switching symbols and terminating the partnership himself. Considering therefore the incentives in the two latter cases, we observe that an s^i player will not choose to unilaterally switch symbol (to s^j, say) if[19]:

$$v^i\left(\bar{e}^i\right) \geq \pi\left(0, \bar{e}^i\right) + \delta w^j\left(\bar{e}^j\right) - c_t\left(i, j\right). \tag{10}$$

Substituting (6) satisfied with equality into (10), we obtain the following characterization of unilateral symbol switching.

Proposition 2 *In an efficiently segregating PPE, the set* $\mathcal{E}(c)$ *of all symbol frequencies that can be sustained as an efficiently segregating PPE is implicitly defined by the following system of inequalities:*

$$w^j\left(\bar{e}^j\right) - w^i\left(\bar{e}^i\right) \leq \frac{c\left(i, j\right)}{\delta}. \tag{11}$$

In particular, these inequalities imply that no player bearing symbol s^i *wants to switch to another symbol* s^j*.*

The set of inequalities present in (11) highlight the importance of the switching costs in determining the set of symbol frequencies compatible with the efficient segregating PPE. Indeed, notice that as the values $c(i, j)$ increase, then the inequalities (11) become less stringent and therefore, the set of feasible symbol frequencies, α^i increases, more formally, if $c'(i, j) \geq c(i, j)$ for all $i \neq j$ and the inequality is strict for at least one pair i, j, then $\mathcal{E}(c) \subset \mathcal{E}(c')$. In particular, it follows that if the switching costs grow indefinitely, then $\mathcal{E}(c) = \mathcal{U}$.

Besides the switching costs, the set $\mathcal{E}(c)$ also depends on the shape of w, the expected continuation values when beginning a partnership with a randomly drawn new partner. Using (4), (5) and also remarking that (6) holds with equality in efficiently segregating PPE, we can write

$$w^i\left(\bar{e}^i\right) = \frac{p^i \pi\left(0, \bar{e}^i\right)}{1 - \delta}. \tag{12}$$

The following Lemma characterizes the shape of $w^i(\bar{e}^i)$ as a function of p^i.

Lemma 2 *In the efficiently segregating PPE,* $w^i(\bar{e}^i)$ *decreases at any discontinuity and is such that*

$$\lim_{p^i \to 0} w^i\left(\bar{e}^i\right) = \lim_{p^i \to 1} w^i\left(\bar{e}^i\right) = 0.$$

Moreover, $w^i(\bar{e}^i)$ *increases with* p^i*, where* $w^i(\bar{e}^i)$ *is differentiable w.r.t.* p^i*, iff* $\frac{\partial d\left(\bar{e}^i, 0\right)}{\partial \bar{e}^i} \leq 0$.

etc.

[19]Notice that, for ease of exposition, we drop the dependency of $\bar{e}^i$ on p^i, λ and δ.

Hence, the matrix of symbol switching costs imposes a bound on the maximal difference in continuation values with a randomly drawn partner. The continuation value of s^i players with a new randomly drawn partner approaches zero for very asymmetric distributions over symbols, α. Such extremely skewed frequencies on the population level translate in extreme values of p^i. Notice that, for $p^i \to 1$, the almost certainty of finding a new s^i partner in the next round prevents them from committing to significant effort levels. If $p^i \to 0$, then the inability of finding a new s^i partner after a partnership termination drives $w^i(\bar{e}^i)$ to zero, despite s^i players being able to sustain the highest possible effort level in a homogenous partnership, which we denote

$$\tilde{e} \equiv \max\{e | d(e, 0) = 0\}.$$

Hence, $\tilde{e}$ is the effort level that can only be sustained by partners who know they will never again find a cooperative partner after the termination of their present partnership. Starting from $p^i = 0$, it is plausible to see the continuation value of s^i players, w^i, initially increase with p^i, because the decrease in sustainable efforts, $\bar{e}^i$, is initially more than compensated for by an increased likelihood of finding a new s^i partner. In particular, and for future purposes, we define p^* as the highest share such that, for all $p^i \leq p^*$, $w^i(\bar{e}^i)$ increases with p^i. Let e^* denote the corresponding effort, such that $d(e^*, p^*) = 0$.

Summarizing, $\mathcal{E}(c)$ consists of all distributions α such that efforts for players are given by (9), continuation values and switching costs are such that no player has incentive to change symbol, i.e. (11) is satisfied. Before turning to a further discussion of the model, it first remains to show that at least one efficient segregating PPE exists. This is done in the following Proposition.

Proposition 3 *For all matrices of symbol switching costs c,, $\mathcal{E}(c) \neq \emptyset$, i.e., we can always find a vector $(\alpha^i)_{i=1,\dots,n}$ that can be sustained as an efficiently segregating PPE.*

The proof of Proposition 3 relies on the fact that the effort levels in Proposition 1 are well defined if condition 1 is satisfied, demonstrates the subgame perfection of the efficient segregating PPE, and argues that Proposition 2 is always satisfied for uniform symbol frequencies, since in that case $w^i(\bar{e}^i) = w^j(\bar{e}^j)$.

4 Gradual trust building

In the previous sections, we restricted our attention to a particular class of PPE. Clearly, there are potentially many more PPE which are also 'symbol-blind', in which the contributed efforts are not conditioned on symbols. A first example of such a symbol-blind PPE is one in which players never exert positive effort. Clearly, such a PPE always exists.[20] Second, among the main candidates for a symbol-blind PPE with strictly positive efforts are those which involve a form of gradual trust-building or incubation. In these equilibria, the equilibrium efforts depend on the age of a partnership, denoted $\tau \in \mathbb{N}$. In the early ('dating') rounds of a partnership, equilibrium efforts are low, and the prospect of facing these low continuation values in a new partnership enables partners to sustain high efforts in later rounds. Thus, a symmetric incubation PPE is characterized by a sequence $(e_\tau)_{\tau=0,1,2,\dots}$ which satisfies infinitely many incentive compatibility constraints, such that for all $\tau \geq 0$:[21]

$$\pi\left(e_\tau, e_\tau\right) - \pi\left(0, e_\tau\right) + \sum_{j=1}^{\infty} \left(\delta\left(1-\lambda\right)\right)^j \left(\pi\left(e_{j+\tau}, e_{j+\tau}\right) - \pi\left(e_{j-1}, e_{j-1}\right)\right) \geq 0. \tag{13}$$

The incentive compatibility constraints in (13) require that, for all τ, the expected future benefits of being $\tau+1$ rounds further advanced in a partnership exceed the benefits of shirking in the τ-th round (and having to start 'trust-building' anew). In addition, players only terminate a partnership if their partner deviates from the equilibrium play. Efficiency then implies maximizing the stream of payofs that players get from a partnership:

$$\sum_{\tau=0}^{\infty} \left(\delta\left(1-\lambda\right)\right)^\tau \pi\left(e_\tau, e_\tau\right), \tag{14}$$

subject to (13) . This characterization illustrates that finding gradual symbol-blind equilibria can be quite tedious in the present public goods game setting with a continuum action space, and that finding efficient equilibria will be very difficult. We will therefore focus our attention to the most simple type of gradual-trust building strategies, namely those in which there is an incubation phase of length T and then a cooperation phase. We will refer to these strategies as T-GTB, formally defined as follows:

Definition 2 (T-GTB - T period Gradual Trust Building) *The class of T-GTB PPE are the stationary PPE in pure strategies in which players:*

- *exert efforts e_0 for the first T periods,*
- *after T periods they exert efforts $e_1 > e_0$,*
- *the partnership is immediately terminated if one of the players deviates from the prescribed effort levels or due to exogenous break up.*

[20]Indeed, this constitutes a Nash equilibrium of the stage game.
[21]The derivation of (13) and (14) is presented in the Appendix.

The strategies involved in this class of (simplified) gradual-trust building PPE are a straightforward extension of the 'incubation strategies' tailored to discrete action spaces (e.g. in prisoner's dilemma games), such as those discussed in Fujiwara-Greve and Okuno-Fujiwara (2009). We begin the analysis of these simple gradual-trust strategies by finding the continuation values. Now, we denote by v_t the (expected) continuation value for a player who has been in a partnership with another player for t periods. Then, we have that

$$v_t = \pi(e_1, e_1) + \delta\left[(1-\lambda) v_{t+1} + \lambda v_0\right] \text{ for } t \geq T,$$

that is, for a partnership that has already lasted (at least) T periods, each player obtains $\pi(e_1, e_1)$. Furthermore, with a probability of $1-\lambda$ the partnership is not broken up, yielding a continuation value of v_{t+1}. With probability λ the relationship is broken up, after which each player draws a new partner and the 'counter' (duration of partnership) is reset to zero, resulting in an expected continuation value of v_0. Notice that, as long as the partnership continues (after reaching together for at least T periods), the continuation values are constant, that is, $v_t = \bar{v}$ for all $t \geq T$. We can then derive an explicit expression for this expected continuation value, in case the partnership survives the incubation phase:

$$\bar{v} = \frac{\pi(e_1, e_1)}{1-\delta(1-\lambda)} + \frac{\delta\lambda}{1-\delta(1-\lambda)} v_0. \tag{15}$$

Continuing with the continuation values for the other periods, we obtain:

$$v_t = \begin{cases} \pi(e_0, e_0) + \delta\left[(1-\lambda)\bar{v} + \lambda v_0\right] \text{ if } t = T-1, \\ \pi(e_0, e_0) + \delta\left[(1-\lambda) v_{t+1} + \lambda v_0\right] \text{ if } t < T-1. \end{cases}$$

Which can be solved to yield the following expression[22]:

$$v_t = \frac{1-(\delta(1-\lambda))^{T-t}}{1-\delta(1-\lambda)} \pi(e_0, e_0) + \delta\lambda \frac{1-(\delta(1-\lambda))^{T-t}}{1-\delta(1-\lambda)} v_0 + (\delta(1-\lambda))^{T-t} \bar{v}. \tag{16}$$

Substituting (15) into (16) then gives:

$$v_t = \frac{(\delta(1-\lambda))^{T-t}}{1-\delta(1-\lambda)} \left(\pi(e_1, e_1) - \pi(e_0, e_0)\right) + \frac{\pi(e_0, e_0) + \delta\lambda v_0}{1-\delta(1-\lambda)}. \tag{17}$$

[22]This can be derived by solving backwards, through repeated substitution of v_{t+1} into v_t.

The effort levels (e_0, e_1) then need to satisfy the following set of incentive compatibility constraints:

$$\pi(e_0, e_0) + \delta[(1-\lambda) v_t + \lambda v_0] \geq \pi(0, e_0) + \delta v_0, \text{ for all } t < T, \tag{18}$$

$$\pi(e_1, e_1) + \delta[(1-\lambda) \bar{v} + \lambda v_0] \geq \pi(0, e_1) + \delta v_0, \text{ for } t \geq T. \tag{19}$$

At this point, we can easily show that we can reduce the set of incentive compatibility constraints in (18) by checking that no player has an incentive to deviate from the gradual-trust strategy at the start of a partnership, that is,

$$\pi(e_0, e_0) + \delta[(1-\lambda) v_t + \lambda v_0] \geq \pi(0, e_0) + \delta v_0,$$

which, after rearranging terms, yields the following:

$$\pi(e_0, e_0) + \delta(1-\lambda) v_0 \geq \pi(0, e_0) + \delta(1-\lambda) v_0.$$

Given our assumptions on the payoff function, cfr. Condition 1, this incentive compatibility constraint can only be satisfied by setting $e_0 = 0$. Given we are focussing on efficient effort levels within the partnership, the effort level post-incubation, e_1 are determined by the following:

$$\max_{e_1} \sum_{\tau \geq T} (\delta(1-\lambda))^\tau \pi(e_1, e_1),$$

subject to (19). After substituting (15) and rearranging terms, we can rewrite (19), in analogy to equation (8) for the efficiently segregating PPE, by defining the following map

$$\tilde{d}(e, T) \equiv \frac{1 - (\delta(1-\lambda))^{T+1}}{1 - \delta(1-\lambda)} \pi(e_1, e_1) - \pi(0, e_1) \geq 0. \tag{20}$$

The value $\tilde{d}(e, T)$ gives the difference between the expected value of the current partnership (with a length of incubation phase given by T and efforts in the cooperation phase given by e) and the one shot payoff of cheating. Efficiency then implies that efforts will be chosen in such a way that the inequality (20) is exhausted. The following summarizes the properties of efforts in a T-GTB equilibrium:

Proposition 4 *In a T-GTB equilibrium, efforts e_1^{GTB} uniquely solve*

$$e_1^{GTB} = \max\left\{e | \tilde{d}(e, T) = 0.\right\}. \tag{21}$$

Moreover, e_1^{GTB} is strictly increasing function of T, right-continuous and strictly increasing with δ and a left-continuous and decreasing function of λ.

The fact that e_1^{GTB} is increasing with T motivates the name 'gradual-trust building', indeed, the longer waiting time (incurring zero payoff) within a partnership allows for more trust and hence higher efforts in the 'cooperation phase', the second phase of the partnership.

To illustrate we can again resort to our closed form example:

Example 3 *For our closed form example, we had $\pi(e_1, e_1) = \frac{e_1}{1+e_1}$ and $\pi(0, e_1) = e_1$, so from this we can deduce that*

$$e_1^{GTB} = \max\left\{e \mid \frac{1-(\delta(1-\lambda))^{T+1}}{1-\delta(1-\lambda)}\frac{e_1}{1+e_1} = e_1.\right\}$$

$$= \frac{\delta(1-\lambda)\left(1-(\delta(1-\lambda))^{T+1}\right)}{1-\delta(1-\lambda)}.$$

This can also be illustrated graphically:

Figure 2: Gradual trust building: efforts with different T

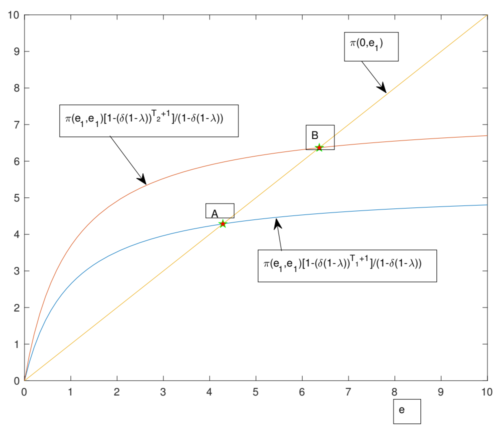


Figure 2 illustrates the dependence of e_1^{GTB} on the length of the 'incubation phase', T. Indeed, we have shown here two cases, one where $T = T_1$ and another with $T = T_2 > T_1$. In the former case, the equilibrium efforts players will pick in the cooperation phase can be read from the intersection point A, whereas B shows the efforts for the longer incubation phase.

4.1 Optimal GTB

Substituting the equilibrium effort levels e_1^{GTB} (as defined in (21)) into the expression for the expected payoffs yields:

$$W_{GTB}(T;\delta,\lambda) = \frac{(\delta(1-\lambda))^T}{1-\delta(1-\lambda)}\pi\left(e_1^{GTB}, e_1^{GTB}\right). \tag{22}$$

From this expression, it is clear that the expected payoffs are non-monotonic with respect to T. Indeed, an increase in T improves expected payoffs in the cooperation phase through higher sustainable efforts, cfr. the expression in (21), but players have to incur a longer incubation phase, in which they get no payoffs. Given this trade-off, we are interested in the optimal length of the incubation phase, i.e., the T that maximizes the expected payoffs of players under the GTB equilibrium. Formally[23],

$$T^* = \arg\max_T W_{GTB}(T;\delta,\lambda). \tag{23}$$

In terms of terminology, we will use the following:

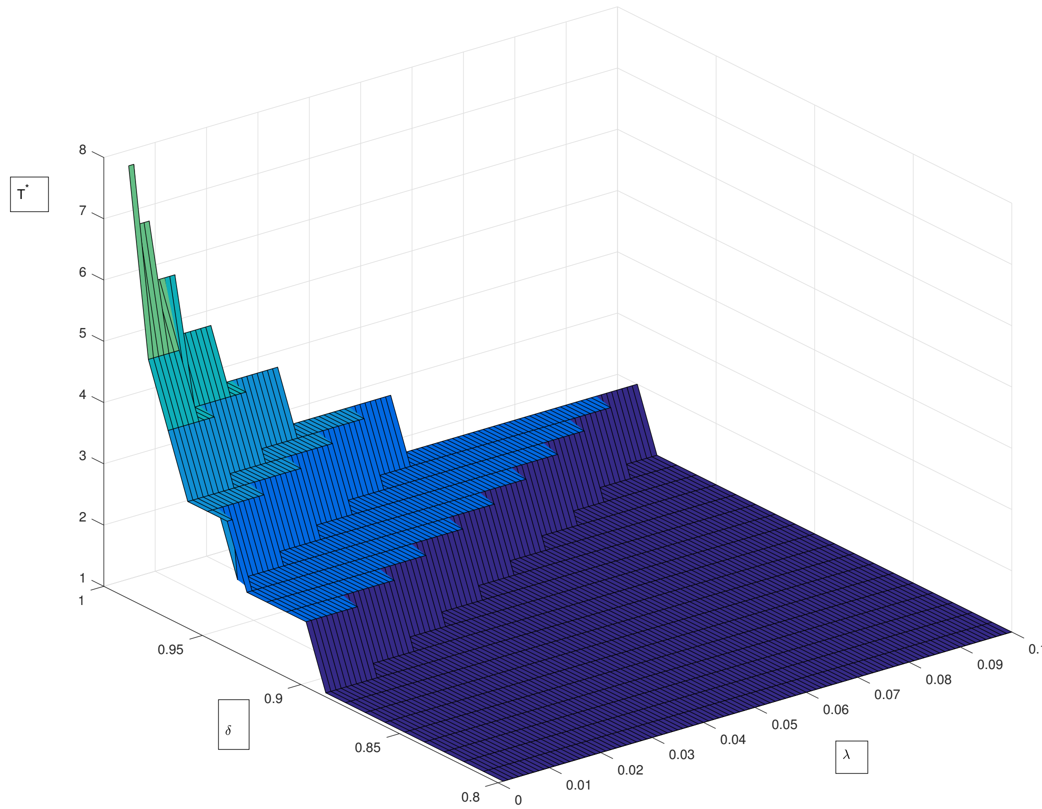
Definition 3 **($T^* - GTB$)** *The $T^* - GTB$ is a T-GTB with length of the incubation phase given by $T = T^*$, as defined in (23).*

Notice that the $T^* - GTB$ is defined conditional on the values for δ and λ. To illustrate how the length of the incubation phase in the $T^* - GTB$ depends on the latter, we can again resort to our running example:

Example 4 *We substitute the expression we derived earlier for $e_1^{GTB} = \frac{\delta(1-\lambda)\left(1-(\delta(1-\lambda))^{1+T}\right)}{1-\delta(1-\lambda)}$ in (22) and then look for the value T^* which maximizes these expected payoffs. We can illustrate the solution T^* as a function of δ and λ,*

[23] Notice that $\lim_{T\to 0^+} W_{GTB}(T;\delta,\lambda) = \lim_{T\to\infty} W_{GTB}(T;\delta,\lambda) = 0$, which implies that T^* is interior.

Figure 3: T^* as function of δ, λ



Larger values of λ and smaller values of δ (hence low effective discount factors $\delta(1-\lambda)$) are associated with shorter optimal incubation phases.

Figure 3 also illustrates that T^* will generally take on the form of a step-function, in particular, given the integer restriction $T^* \in \mathbb{N}_0$ on the range of (23), if we fix a value of λ, then for any value of δ there will be a range of discount factors that give the same value of T^*. Similarly, if we fix a value of δ, then for all λ there will be a range around this exogenous probability of break up that gives rise to the same optimal length of the incubation phase.[24]

[24]More formally, if we fix a value for λ, say $\bar{\lambda} \in (0,1)$, then for all $\bar{\delta}$, there will exist a $k > 0$ such that $T^*\left(\delta, \bar{\lambda}\right) = T^*\left(\bar{\delta}, \bar{\lambda}\right)$, for all $\delta \in \left(\bar{\delta} - k, \bar{\delta} + k\right)$. Similarly, if we fix a value $\tilde{\delta} \in (0,1)$, then for all $\tilde{\lambda} \in (0,1)$, there exists a $k > 0$ such that $T^*\left(\tilde{\delta}, \lambda\right) = T^*\left(\tilde{\delta}, \tilde{\lambda}\right)$.

4.2 Comparing gradual trust-building with efficiently segregating PPE

We now want to compare the optimal gradual trust-building equilibria with the efficiently segregating PPE. To be more precise, we will compare the (expected) payoffs for the gradual trust-building equilibria with the average expected payoffs under efficient segregation. To make progress, we can compute the average expected payoffs under an efficiently segregating PPE. In particular, for a player bearing symbol s^i and a distribution of symbols α this is given by[25]:

$$W^i(\alpha; n) = \left(\alpha^i - x^i\right) v^i\left(\bar{e}^i\right) + x^i w^i\left(\bar{e}^i\right). \tag{24}$$

A player bearing symbol s^i spends on average a fraction x^i of her time having to draw a new partner, yielding an (expected) continuation value $w^i(\bar{e}^i)$ and a fraction $\alpha^i - x^i$ in a homogeneous partnership which gives a continuation value of $v^i(\bar{e}^i)$, combining these gives the expression for the average expected value in (24). A natural way to compare equilibria is in terms of Pareto dominance, which we will also employ here. The following summarizes our first result:

Proposition 5 *For every gradual trust-building PPE with an associated length of the incubation phase given by $T > 0$, we can find an efficiently segregating PPE, with symbols' distribution $\alpha \in \mathcal{E}(c)$, such that*

$$W^i(\alpha; n) \geq W_{GTB}(T; \delta, \lambda), \text{ for all } s^i, \tag{25}$$

as long as $\delta, \lambda \in (0,1)$. Furthermore, $W^i(\alpha; n) - W_{GTB}(T; \delta, \lambda)$ is decreasing with δ and increasing with λ.

According to Proposition 5 we can always find, for every gradual-trust building strategy, an efficiently segregating PPE which yields higher (or not worse) average expected payoffs for all players. In order to provide more insight in the result, we provide a brief sketch of the proof here. We search for a uniform distribution of symbols $\alpha = \left(\frac{1}{n}, \ldots \frac{1}{n}\right)$, which is chosen in such a way such that the effort levels in the corresponding efficiently segregating PPE equal those of the GTB PPE (given a length of the incubation phase given by T.) [26] The

[25] And we assume $\alpha \in \mathcal{E}(c)$, i.e., there is no symbol switching.

[26] It is important to notice that we require S to be sufficiently 'rich' (i.e. large n). Indeed, it can be shown that equality of efforts with a uniform distribution over symbols implies $n = \frac{1-(\delta(1-\lambda))^{T+1}}{(\delta(1-\lambda))^T[1-\delta(1-\lambda)]}$. Note that since $n \in \mathbb{N}_0$, we select $\lceil \frac{1-(\delta(1-\lambda))^{T+1}}{(\delta(1-\lambda))^T[1-\delta(1-\lambda)]} \rceil$, where we recall that, for $z \in \mathbb{R}$, $\lceil z \rceil$ denotes the least integer greater than or equal to z.

remainder of the proof then boils down to directly compare the resulting (average) expected payoffs from the corresponding efficiently segregating PPE and the GTB. Notice that the Proposition excludes those non-generic cases in which $\delta = 0$ or $\lambda = 1$, since these cover the non-interesting setting in which no player exerts any effort. Furthermore, Proposition 5 states that the relative benefits of the (dominating) efficiently segregating PPE' are reduced for larger values of the effective discount factors for players, i.e. larger values of $\delta(1-\lambda)$. Intuitively, more patient players can afford to incur low payoffs in the incubation phase and thereby receive larger expected payoffs in the cooperative phase of the T-GTB equilibrium. This mechanism thereby reduces the relative advantage of the efficiently segregating PPE over the T-GTB PPE.

We are now interested in whether we can extend the result in Proposition 5 to the optimal GTB, i.e. those gradual trust-building PPE in which the length of the incubation phase is given by T^*, as defined by (23). The main complication in this case is that the length of the incubation phase automatically (and optimally) adjusts in response to different environments, that is, to different values for δ and λ, which makes direct comparison between the best GTB and the class of efficiently segregating PPE more difficult. However, we are able to show that the result in Proposition 5 still holds:

Proposition 6 *Consider a given $T^* - GTB$. We can always find an efficiently segregating PPE with initial symbols' distribution $\alpha \in \mathcal{E}(c)$ such that:*

$$W^i(\alpha; n) \geq W_{GTB}(T^*(\delta, \lambda); \delta, \lambda), \text{ for all } s^i, \tag{26}$$

as long as $\delta, \lambda \in (0,1)$. The difference $W^i(\alpha; n) - W_{GTB}(T^*; \delta, \lambda)$ is decreasing with δ and increasing with λ.

The result in Proposition 6 has essentially the same content as in Proposition 5. The results in Propositions 5 and 6 motivates our focus on the class of efficiently segregating PPE in the present paper, in the sense that one can always find an efficiently segregating PPE which Pareto dominates any given gradual trust-building PPE. Also notice that, even though the proof of this result uses a uniform symbols' distribution, the result itself doesn't preclude that non-uniform symbol distributions can give Pareto dominant (average) expected payoffs to all players. Indeed, moving probability mass in α from one symbol, say s^j to another symbol, s^i, would decrease (by the enveloppe theorem) the average expected payoffs W^j for the s^j-bearing players. However, if (25) holds with strict inequalities, then the new symbols' distribution would still yield average expected payoffs which are Pareto dominating those for the T-GTB PPE. We can repeat this exercise with an arbitrary combination of symbols, until

(25) is satisfied with equality for at least one symbol. Note that the symbol switching costs c play an important role in these iterations, since every newly obtained distribution over symbols should yield continuation values that satisfy (11), i.e., no player would be willing to switch symbols and pay the cost of doing so.

5 Optimality

So far we characterized the efficiently segregating PPE *given* a particular set of symbols, S and for an initial distribution of the players over symbols, α. In this section, we are interested in comparative static results on how changes in S and the initial distribution α affect (a measure for) the overall payoffs (i.e., average across all symbols). In particular, we will focus on the following measure:

$$W(\alpha; n) = \sum_{i=1}^{n} W^i(\alpha; n) = \sum_{i=1}^{n} \left(x^i w^i\left(\bar{e}^i\right) + \left(\alpha^i - x^i\right) v^i\left(\bar{e}^i\right)\right). \tag{27}$$

Some remarks are in order. First, note that, given the distribution of symbols α, one can compute equilibrium efforts $\bar{e}^i$ in a corresponding efficiently segregating PPE, for all s^i via Proposition 1. The latter then pin down the (expected) continuation values in homogeneous and heterogeneous partnerships[27]. Hence, we simply write W as a function of α. Second, notice that (27) is simply the sum of the average expected payoffs across all the symbols, the latter given by (24).

5.1 The uniform case

We start the analysis of how W is affected by S using the uniform benchmark, i.e., we assume that $\alpha = \left(\frac{1}{n}, ..., \frac{1}{n}\right)$. In this case, the continuation values $w^i\left(\bar{e}^i\right) = w^j\left(\bar{e}^j\right) = w\left(\bar{e}\right)$ and therefore, (11) is automatically satisfied. Furthermore, it can be easily shown using (2) that for the uniform case, $x^i = x$ for all s^i, and hence $p^i = \frac{1}{n}$, for all s^i. Furthermore, making use of the fact that (6) holds with equality in an efficiently segregating PPE, together with (12), we can derive the following expression:

$$W\left(\left(\frac{1}{n}, ..., \frac{1}{n}\right); n\right) = \left[(1 - nx)\left(1 + \frac{\delta}{n(1-\delta)}\right) + \frac{x}{1-\delta}\right] \pi\left(0, \bar{e}^i\right). \tag{28}$$

From (28) it is clear that increasing the number of symbols, n, has a non-monotonic effect on the overall average payoff for all players. Indeed, under the assumption of a uniform

[27] As usual, under the assumption that there is no symbol switching in equilibrium, i.e., under the assumption that $\alpha \in \mathcal{E}(c)$ and therefore (11) is satisfied.

distribution of players across symbols, more symbols implies smaller groups which, following Proposition 1 means that players will be able to sustain higher levels of efforts within homogeneous partnerships. However, when the number of symbols becomes too large, each player will spend more time searching for a new partner, each time incurring the lower continuation value $w(\bar{e})$. We can illustrate this non-monotonic feature more clearly in the context of our running example:

Example 5 *First, for our running example, the expression for the equilibrium efforts in the uniform case is given by,*

$$\bar{e} = \frac{(n-1)\,\delta\,(1-\lambda)}{n-(n-1)\,\delta\,(1-\lambda)},$$

which follows after substituting $p^i = \frac{1}{n}$ *in the expression for* $\bar{e}^i$ *(cfr. supra.) Also recall that* $\pi(0,\bar{e}) = \bar{e}$*. From this, it easily follows that*

$$W\left(\left(\frac{1}{n},...,\frac{1}{n}\right);n\right) = \left[(1-nx)\left(1+\frac{\delta}{n\,(1-\delta)}\right)+\frac{x}{1-\delta}\right]\frac{(n-1)\,\delta\,(1-\lambda)}{n-(n-1)\,\delta\,(1-\lambda)}.$$

The relationship between W *and* n *can now easily be represented graphically,*

Figure 4: Number of symbols and aggregate average payoffs

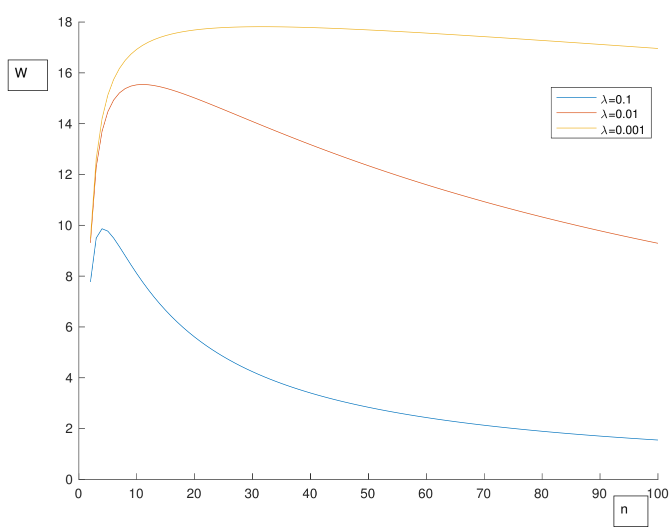


Figure 4 not only highlights the non-monotonicity of W with respect to the number of symbols, but it also shows how the exogenous break-up probability λ impacts the optimal number of symbols (that is, the number of symbols such that W is maximized, under the assumption of a uniform distribution α.) In particular, lower values for the (exogenous) break-up probability for a partnership increase this optimal number of symbols. The logic behind this result comes from the fact that lower values for λ and thus higher effective discount factor, $\delta(1-\lambda)$ imply higher sustainable efforts in the efficiently segregating PPE. This increases the continuation values as a first order effect. A natural question which arises is whether (27) is maximized by the uniform distribution, $\alpha = \left(\frac{1}{n}, ..., \frac{1}{n}\right)$. The answer is in general negative.To see this, first for given δ and λ, let

$$n^* = \arg\max_n W\left(\left(\frac{1}{n}, ..., \frac{1}{n}\right); n\right), \tag{29}$$

i.e., the number of symbols for which the average expected payoffs (across all symbols) is maximized under the assumption that α is uniform. Also recall that $W\left(\left(\frac{1}{n}, ..., \frac{1}{n}\right); n\right)$ is given by (28). We can then state the result:

Proposition 7 *If the number of symbols, $n > n^*$, then there exists a distribution $\alpha \in \mathcal{E}(c)$ with $\alpha \neq \left(\frac{1}{n}, ..., \frac{1}{n}\right)$ for which*

$$W(\alpha; n) > W(1/n, ..., 1/n; n). \tag{30}$$

The idea for the proof is straightforward. Given that the number of symbols exceeds the critical value n^*, there exists a subset of symbols for which a uniform distribution over this subset yields a higher overall (average) expected payoff than the uniform distribution over the entire set of symbols. Hence, starting from our initial distribution $\left(\frac{1}{n}, ..., \frac{1}{n}\right)$, we can find a subset of symbols, say $\tilde{n}$ of them, where $n^* \leq \tilde{n} < n$, such that

$$\alpha = \left(\underbrace{1/n+\epsilon, ..., 1/n+\epsilon}_{\text{length } \tilde{n}}, \overbrace{1/n-\epsilon, ..., 1/n-\epsilon}^{\text{length } n-\tilde{n}}\right),$$

and it can be easily seen that (30) will be satisfied for a sufficiently small value for $\epsilon > 0$.

6 Bilateral renegotiation

So far we have focussed on robustness of equilibria with respect to unilateral deviations. Though this is a standard approach in analyzing these sort of games, it does leave open some unappealing and counterintuitive equilibria. For example, consider a PPE in which there is no cooperation, in particular a case whereby no player ever exerts (positive) efforts in a partnership with another player. It is clear that this is a PPE, as no single player has an incentive to unilaterally deviate to exert a positive amount of effort while his partner sticks to the equilibrium zero effort strategy. But if the two current partners can mutually improve themselves by jointly deviating to strictly positive efforts, and if such a joint deviation is incentive compatible, then we expect both players to take advantage of it. Such a refinement of 'bilateral rationality' is overly strong, indeed, it eliminates all PPE (Kranton (1996), Ghosh and Ray (1996)). This then also carries through to our class of efficiently segregating PPE. However, not all PPE might be equally sensitive to 'some' form of bilateral deviation. More formally, we will quantify the'size' of bilateral deviations for a particular (class of) PPE to survive. This allows us to compare different (classes of) PPE in terms of these bounds to bilateral renegotiations.

To make progress, we need to have a measure for the size of deviations. We opt for a definition of distance on the strategy space, in particular for two strategies σ and σ' let $m(.)$ represent, for the sake of simplicity, a continuously differentiable function that strictly increases with respect to the differences in efforts, measured by the Euclidian metric, and differences in termination decisions and symbol switching, which are both measured by a discrete metric.[28] The distance between σ and σ' is then given by

$$M(\sigma, \sigma') = \sup_{h_t} m(\sigma(h_t), \sigma'(h_t)) \tag{31}$$

Our first result with respect to bilateral deviations shows that the gradual trust-building PPE are vulnerable to *any* sort of bilateral deviation, no matter how small:

Proposition 8 *Let σ be an equilibrium strategy for a player consistent with a GTB PPE with associated length for the incubation phase given by T. Then, for all $\varepsilon > 0$ there exists a jointly profitable (bilaterally agreed) deviating strategy σ' with $M(\sigma, \sigma') \leq \varepsilon$.*

The result in Proposition 8 states that, for a T-GTB type equilibrium with a positive length of the incubation phase ($T > 0$), one can always find a jointly profitable deviating strategy. To provide some intuition: note that (regarding what players can expect from a new partnership) the equilibrium effort sequence $\left(\underbrace{0, ..., 0}_{\text{length} = T}, e_1, ..., e_1\right)$ fixes the same outside option for all players, independent of the age of their current partnership. If this outside option makes high efforts enforceable in later rounds of a partnership, then these high efforts are equally enforceable in the first round. As long as all others play the equilibrium strategies, two current partners can mutually improve themselves by jointly deviating to higher efforts in the first round of their partnership, up to the point where they exert the highest sustainable

[28]The particular choices for the component-distance functions can be generalized and our results don't fundamentally rely on these choices.

efforts. And such a joint deviation is profitable for both partners even if we only allow for a very small deviation. Hence, robustness against even the smallest joint deviations requires that the equilibrium efforts are independent of the partnership's age. If the constant equilibrium efforts are high, then they violate incentive compatibility. If they are low, then they constitute a bad outside option, giving rise to bilateral deviations to higher efforts. From this reasoning, it follows that the only case in which a PPE that is 'symbol blind' survives bilateral renegotiation of any size is when $\delta(1-\lambda) = 0$, that is, when players effectively play a sequence of one shot public goods games.

We now turn to the class of efficiently segregating PPE. In this case, there are two possible joint deviations: first, the effort levels within homogeneous partnerships can be renegotiated, hence these should be sufficiently high so as to avoid that low equilibrium efforts allow two same symbol players to credibly commit to a joint deviation to higher efforts. Second, players might be tempted to break the segregation required in the efficiently segregating PPE by jointly deviating to cooperation in a heterogeneous partnership, in order to avoid waiting for a same symbol partner. Now, suppose that within a homogeneous partnership the players agree to exert higher efforts, to the point of e, then from (4), (5) and (6), and keeping $w^i(\bar{e}^i)$ fixed at the efficient segregating PPE level, we have that such a joint deviation is viable if

$$d(e,0) = \frac{\pi(e,e)}{1-\delta(1-\lambda)} - \pi(0,e) \geq \delta\frac{(1-\delta)(1-\lambda)}{1-\delta(1-\lambda)} w^i\left(\bar{e}^i\right), \tag{32}$$

where we recall that $d(e,0)$ is given by (8). Now notice that (32) is satisfied with equality for $e = \bar{e}^i$. By Lemma 2, the left hand side of (32) is decreasing with e for $e \geq e^*$, such that robustness to joint deviations in homogeneous partnerships requires that p^i is smaller than p^* for all s^i. With respect to the second type of deviations, let σ' be a bilaterally agreed deviation. Then note that if we fix $\varepsilon = M(\sigma, \sigma')$, we can define the maximum deviation in terms of efforts[29] as a function of ε, $\hat{e}(\varepsilon)$. Now, consider the case of a pair of s^i and s^j players, and assume without loss of generality that $p^i \leq p^j < p^*$. Clearly, joint deviations to efforts above $\bar{e}^j$ are not credible, because they fail to satisfy incentive compatibility for the s^j player. The s^i player should thus settle for less efforts than his equilibrium efforts in homogeneous partnerships but can nevertheless be tempted to avoid waiting for a homogeneous partnership. It can be shown that a sufficient condition for a

[29] Furthermore, note that in this case there is a further deviation in the equilibrium strategy in terms of termination decision, which should be taken into account when deriving the maximal effort levels in a bilateral deviation.

joint deviation to $\hat{e}(\varepsilon)$ to not be viable reads: [30]

$$\frac{\pi\left(\hat{e}(\varepsilon),\hat{e}(\varepsilon)\right)}{1-\delta(1-\lambda)}-\pi\left(0,\hat{e}(\varepsilon)\right)<\frac{\delta(1-\lambda)(1-\delta)}{1-\delta(1-\lambda)}w^{i}\left(\bar{e}^{i}\right). \tag{33}$$

The inequality in (33) effectively bounds the maximal deviations in terms of efforts. Furthermore, given that we are interested in the maximum amount of bilateral deviation strategies for which the efficiently segregating PPE survives, we replace the inequality in (33) by an equality. The latter then gives the largest possible value for $\hat{e}(\varepsilon)$ and consequently, the largest possible ε, i.e. range of deviations in strategy space to which the efficiently segregating PPE is robust against. Denoting by $\mathcal{N}(\sigma;\varepsilon)$ the set of all strategies σ' which satisfy $M(\sigma,\sigma')\leq\varepsilon$, we can summarize the result as follows:

Proposition 9 *Let σ be the equilibrium strategy associated with a particular efficiently segregating PPE, with corresponding symbols' distribution $\alpha\in\mathcal{E}(c)$. Then there exists a $\varepsilon>0$ such that σ is robust to any bilateral deviations $\sigma'\in\mathcal{N}(\sigma;\varepsilon)$.*

The result in Proposition 9 essentially provides another motivation to study the class of efficiently segregating PPE, since it essentially postulates that this class of PPE are more robust to (some amount of) 'bilateral rationality' compared to the class of symbol-blind PPE, such as those of the GTB type. Of course, the principal reason that the efficient segregated PPE is robust with respect to sufficiently small joint deviations by current partners, unlike symbol-blind PPE, lies in the radically different equilibrium efforts in homogeneous and heterogeneous partnerships. Restrictions on the size of joint deviations constrain in particular the profitability of deviations to cooperation in heterogeneous partnerships. As such, the restriction to small joint deviations can be understood as a reduced form of a model where other factors impede joint deviations in heterogeneous partnerships, and unbounded renegotiation proofness can thus be obtained by, e.g., permutations in the players preferences.[31] One example of such impediments are chauvinist or intolerant preferences, as modelled by Corneo and Jeanne (2009, 2010). These authors study the evolution of intolerance, i.e., how the appreciation for diversity ('tolerance') evolves, by considering how parents socialize their children by dividing a unit of 'symbolic' valuation over different types people, e.g., different professions, sexual orientations or ethnicities. Their offspring's later evaluation of their own

[30] See the proof of Proposition 9 in the Appendix.

[31] In a related setting, Choy (2018) shows how reputational concerns can discourage joint deviations to heterogeneous cooperation in a repeated public goods game setting with endogenous partnership termination, in which players observe the group affiliations of their partners' previous partners. Choy characterizes an equilibrium in which identifiable groups are hierarchically ordered and members of higher groups protect their reputation by refusing any contact with members of lower groups, which in turn ensures a sufficiently low outside option to sustain cooperation.

life then depends on how this distribution of valuations matches their own profession, sexual orientation or frequent interaction with other ethnicities. Corneo and Jeanne show that parents choose a narrow distribution of valuations (intolerance) in static and predictable environments. This maximizes their offspring's expected valuation of their own life, but can also induce them to avoid certain professional choices or interactions at the cost of foregoing economic opportunities. In dynamic environments, parents choose to raise their offspring in tolerance.

7 Conclusion

In this paper we analyzed how segregation in large anonymous societies allows the members from each of the different groups to enforce cooperative behavior within each group. To do this, we made use of the formal framework of repeated games with (anonymous) random matching and introduced and analyzed a particular class of stationary PPE, the efficiently segregating PPE. In these equilibria, players only exert (positive) effort in partnerships with players bearing the same symbol, while in partnerships with players bearing a different symbol they exert zero efforts. Given that the latter are a waste of time, players will also immediately break up and search for a new partner who bears the same symbol. We have shown that players from smaller symbol groups, given the fact that they face on average longer waiting times, will be able to sustain higher efforts in partnerships with each other. This confirms the earlier insights explored in Iannaccone (1992), however, whereas the latter assumes an exogenous negative response to out-group symbols, we derived such behavior endogenously in equilibrium. Furthermore, even though segregation has been studied previously in the literature, we analyzed the effects of segregation on efforts in a continuous action space and where the symbols (hence the groups) are taken as a choice determined as a part of equilibrium behavior. This marks a difference with the contributions from Eeckhout (2006), who studies cooperative behavior in the context of exogenously fixed group identities and for prisoner's dilemma games, and Choy (2018), who studies cooperative behavior using an exogenously given hierarchy of group identities (e.g. a caste system).

We note that although players do not switch symbol in equilibrium, the matrix of symbol switching costs fundamentally determines the group sizes and effort levels that can be sustained in equilibrium. Historically, the majority of the population in the Western world lived in relatively small scale communities. These communities were often distrustful of outsiders, implying high costs of switching to another group, and were thus able to sustain order by means of strong social norms (which, as a restriction of individual self-interest, can be interpreted as a practical manifestation of the effort levels in the present paper). Economic development and integration as well as advances in transportation and communication technology decreased impediments on mobility of contacts across individuals of different groups. Voltaire (2002) famously applauded in his Letters Concerning the English Nation the ability of trade and economic interactions to break down walls between communities, and thus fostering greater freedom:

> "Take a view of the Royal Exchange in London, a place more venerable than many courts of justice, where the representatives of all nations meet for the benefit of mankind. There the Jew, the Mahometan, and the Christian transact together, as though they all professed the same religion, and give the name of infidel to none but bankrupts."

Almost three centuries later, many cheer the internet and its anonimity (and ease of adopting new online identities) for promoting freedom, while many others deplore the degradation of courtesy and good manners that appears to proceed from it. Today's large-scale urban societies are less fragmented in small geographically defined groups, but display instead an impressive cultural and subcultural diversity. The above analysis suggests that these (sub-)cultural identities can be helpful to sustain cooperation and social norms, and that groups that are in need of strong commitment to the common cause need to find group symbols that are very costly to get rid off, e.g. prison tattoos in the context of mutual protection or crime (Gambetta, 2011).

In the second half of the paper, we compared the class of efficiently segregating PPE with symbol-blind strategies to obtain cooperative behavior In particular, the so-called gradual trust-building strategies (Datta (1996), Kranton (1996), Fujiwara-Greve and Okuno-Fujiwara (2009) and Fujiwara-Greve et al. (2012)). In these, players build 'trust' by starting at low (or zero) effort levels and then increasing it to a higher level. Though appealing, we have shown that, for every such gradual trust PPE we can find a Pareto dominating efficiently segregating PPE. Finally, we have shown that, even though all PPE in the game environment we are studying are not robust against bilateral deviations, i.e., 'bilateral rationality'-type refinements as presented in Kranton (1996), Ghosh and Ray (1996), the class of efficiently segregating PPE are robust to allowing players to jointly deviate (within partnerships) in a neighborhood around the equilibrium strategies. This is in contrast to the T-GTB, where players will always want to jointly deviate from the corresponding equilibrium strategies, no matter how small-size deviations are allowed for. These arguments reconfirm a strong motivation to study and analyze the class of efficiently segregating PPE.

Finally, we also provided results in terms of how average payoffs of players are affected by the number of symbols in society and we have shown that in general the uniform distribution of players across symbols is not necessarily optimal for the average payoff of players across all symbols.

References

Aimone, J., L. Iannaconne, M. Makowski, and J. Rubin (2013). Endogenous group formation via unproductive costs. *Review of Economic Studies 80*, 1215–1236.

Berman, E. (2000). Sect, subsidy, and sacrifice: An economist's view of ultra-orthodox jews. *Quarterly Journal of Economics 115*, 905–953.

Berndt, C. E. (2007). Hostile territory: Hightension religion and the jewish peddler. *The American Journal of Economics and Sociology 66*(5), 1005–1027.

Bhaskar, V. and C. Thomas (2019). Observations on cooperation. *Review of Economic Studies 86*(3), 1010–1032.

Choy, J. P. (2018). Social division with endogenous hierarchy. *The Economic Journal 128*, 2711–2742.

Corneo, G. and O. Jeanne (2009). A theory of tolerance. *Journal of Public Economics 93*, 691–702.

Corneo, G. and O. Jeanne (2010). Symbolic values, occupational choice, and economic development. *European Economic Review 54*(2), 237–251.

Datta, S. (1996). Building trust. Working paper, London School of Economics.

Deb, J. (2020). Cooperation and community responsibility. *Journal of Political Economy 128*(5), 1976–2009.

Deb, J. and J. González-Díaz (2019). Enforcing social norms: Trust-building and community enforcement. *Theoretical Economics 14*(4), 1387–1433.

Eeckhout, J. (2006). Minorities and endogenous segregation. *Review of Economic Studies 73*, 31–53.

Ellison, G. (1994). Cooperation in the prisoner's dilemma with anonymous random matching. *Review of Economic Studies 61*(3), 567–588.

Fujiwara-Greve, T. and M. Okuno-Fujiwara (2009). Voluntarily separable repeated prisoner's dilemma. *Review of Economic Studies 76*(3), 993–1021.

Fujiwara-Greve, T., M. Okuno-Fujiwara, and N. Suzuki (2012). Voluntarily separable repeated prisoner's dilemma with reference letters. *Games and Economic Behavior 74*(2), 504–516.

Gambetta, D. (2011). *Codes of the Underworld: How Criminals Communicate*. Princeton, New Jersey: Princeton University Press.

Ghosh, P. and D. Ray (1996). Cooperation in community interaction without information flows. *Review of Economic Studies 63*(3), 491–519.

Greif, A. (1993). Contract enforceability and economic institutions in early trade: The maghribi traders' coalition. *American Economic Review 83*(3), 525–548.

Heller, Y. and E. Mohlin (2018). Observations on cooperation. *Review of Economic Studies 85*(4), 2253–2282.

Iannaccone, L. R. (1992). Sacrifice and stigma - reducing free-riding in cults, communes, and other collectives. *Journal of Political Economy 100*, 271–291.

Kandori, M. (1992). Social norms and community enforcement. *The Review of Economic Studies 59*(1), 63–80.

Kranton, R. E. (1996). The formation of cooperative relationships. *Journal of Law, Economics & Organization 12*(1), 214–233.

Mailath, G. J., L. Samuelson, and S. Avner (2000). Endogenous inequality in integrated labor markets with two-sided search. *American Economic Review 90*(1), 46–72.

Manzini, P., A. Sadrieh, and N. Vriend (2009). On smiles, winks and handshakes as coordination devices. *The Economic Journal 119*(537), 826–854.

Okuno-Fujiwara, M. and A. Postlewaite (1995). Social norms and random matching games. *Games and Economic Behavior 9*, 79–109.

Peski, M. and B. Szentes (2013). Spontaneous discrimination. *American Economic Review 103*(6), 2412–2436.

Sandler, T. and J. Tschirhart (1997). Club theory: Thirty years later. *Public Choice 93*(3-4), 335–355.

Schumacher, H. (2013). Imitating cooperation and the formation of long-term relationships. *Journal of Economic Theory 148*(1), 409–417.

Shimizu, H. (2011). Social cohesion and self-sacrificing behavior. *Public Choice 149*, 427–440.

Tajfel, H. and J. C. Turner (1979). *The Social Psychology of Intergroup Relations*, Chapter An Integrative Theory of Intergroup Conflict. Monterey: Brooks/Cole.

Takahashi, S. (2010). Community enforcement when players observe partners' past play. *Journal of Economic Theory 145*(1), 42–62.

Voltaire (1731 (2002)). *Letters concerning the English Nation.* Electronic Classics Series. http://www.naturalthinker.net/trl/texts/Voltaire,Francois/Voltaire - Letters on England.pdf: Pennsylvania State University.

Watson, J. (1999). Starting small and renegotiation. *Journal of economic Theory 85*(1), 52–90.

Watson, J. (2002). Starting small and commitment. *Games and Economic Behavior 38*(1), 176–199.

A Proofs and derivations

Proof of Proposition 1

Solving (4) and (5) for $v^i(\bar{e}^i)$ and $w^i(\bar{e}^i)$, one obtains

$$v^i(\bar{e}_i) = \frac{1}{1-\delta}\frac{1-(1-p^i)\delta}{1-\delta(1-p^i)(1-\lambda)}\pi\left(\bar{e}^i,\bar{e}^i\right) \tag{34}$$

and

$$w^i\left(\bar{e}^i\right) = \frac{1}{1-\delta}\frac{p^i}{1-\delta(1-p^i)(1-\lambda)}\pi\left(\bar{e}^i,\bar{e}^i\right). \tag{35}$$

Substituting (34) and (35) into the incentive compatibility constraint in (6) , and noting that efficiency implies that the incentive constraint in (6) is satisfied with equality, we obtain after rearranging terms:

$$d\left(\bar{e}^i,p^i\right) = \frac{\pi(\bar{e}^i,\bar{e}^i)}{1-\delta(1-\lambda)(1-p^i)} - \pi\left(0,\bar{e}^i\right) = 0. \tag{36}$$

Note that this also means that $\bar{e}^i$ solves

$$\frac{\pi(\bar{e}^i,\bar{e}^i)}{\pi(0,\bar{e}^i)} = 1-\delta(1-\lambda)\left(1-p^i\right).$$

Under condition 1, it cannot be excluded that $\frac{\pi(e,e)}{\pi(0,e)}$ strictly increases with e on some intervals of $\mathbb{R}_+$. By efficiency, we select the highest e satisfying (36) by means of the maximum operator in (9).

This characterization of $\bar{e}^i$ is well defined if π satisfies condition 1, as the ratio $\frac{\pi(e,e)}{\pi(0,e)}$ continuously maps $\mathbb{R}_+$ on the entire unit interval. First, continuity is implied by the continuous differentiability of π, and $\frac{\pi(e,e)}{\pi(0,e)} \in [0,1]$ for all e because $\pi_1(.) \leq 0$ and $\pi(e,e) \geq 0$ for all e, and because π_2 is bounded away from zero. Second, $\lim_{e\to 0^+} \frac{\pi(e,e)}{\pi(0,e)} = 1$ by the third part of condition 1. This also means that $\lim_{p^i\to 1^-} \bar{e}^i = 0$. And third, $\lim_{e\to\infty} \frac{\pi(e,e)}{\pi(0,e)} = 0$ by the boundedness of $\pi(e,e)$ and because $\pi_2(.)$ is bounded away from zero.

Finally, note that $\bar{e}^i$ decreases continuously with p^i, except where $\bar{e}^i$ constitutes a local maximum of $\frac{\pi(e,e)}{\pi(0,e)}$. At such point, (9) decreases discontinuously to a lower effort level, such that $\bar{e}^i$ constitutes left-continuous and strictly decreasing function of p^i. In a similar fashion, $\bar{e}^i$ is right-continuous and strictly increasing function of $\delta(1-\lambda)$.

Proof of Proposition 2

An s^i player in a heterogeneous partnership wishes to switch to symbol s^j if

$$c(i,j) + \delta w^j\left(\bar{e}^j\right) > \delta w^i\left(\bar{e}^i\right),$$

while an s^i player in a homogeneous partnership switches to symbol s^j if

$$\pi\left(0,\bar{e}^i\right) + c(i,j) + \delta w^j\left(\bar{e}^j\right) > v^i\left(\bar{e}^i\right).$$

Because (6) is satisfied with equality in equilibrium, both these inequalities are equivalent. In a homogeneous partnership, switching symbol after a defection strictly dominates switching symbol after exerting the equilibrium effort, as argued in the text. Rearranging terms gives the inequality provided in Proposition 2.

Proof of Lemma 2

Using the expected continuation value when matched with a randomly drawn partner, $w^i(\bar{e}^i)$,

$$w^i\left(\bar{e}^i\right) = \frac{p^i\pi\left(0,\bar{e}^i\right)}{1-\delta},$$

and noting that $\pi(0,.)$ is a continuous map (by Condition 1), $w^i(\bar{e}^i)$ is a continuous composition of $\bar{e}^i$, which itself, by Proposition 1, is a left continuous function of p^i. Therefore,

$w^i\left(\bar{e}^i\right)$ is a left-continuous function of p^i. As $\bar{e}^i$ decreases at each discontinuity, and $\pi\left(0,.\right)$ is increasing, $w^i\left(\bar{e}^i\right)$ is decreasing at each discontinuity. Using the characterization of $\bar{e}^i$, we have that:

$$\lim_{p^i\to 0}\pi\left(0,\bar{e}^i\right)=\lim_{p^i\to 0}\frac{\pi\left(\bar{e}^i,\bar{e}^i\right)}{1-\left(1-p^i\right)\delta(1-\lambda)}.$$

And for fixed $\delta(1-\lambda)$ bounded away from 1, the latter limit exists and is bounded, because $\pi\left(e,e\right)$ is bounded for all $e\in\mathbb{R}_+$ (by Condition 1). Therefore,

$$\lim_{p^i\to 0}w^i\left(\bar{e}^i\right)=\frac{1}{1-\delta}\lim_{p^i\to 0}\frac{p^i\pi\left(\bar{e}^i,\bar{e}^i\right)}{1-\left(1-p^i\right)\delta(1-\lambda)}=0.$$

In case $p^i\to 1$, we have that $\bar{e}^i\to 0$ such that $\pi\left(0,\bar{e}^i\right)$ converges to zero, and therefore, $\lim_{p^i\to 1}w^i\left(\bar{e}^i\right)=0$. Consider a p^i at which $w^i\left(\bar{e}^i\right)$ is differentiable. Now, solving for $v^i\left(\bar{e}^i\right)$ in (4) gives us:

$$v^i\left(\bar{e}^i\right)=\frac{\pi\left(\bar{e}^i,\bar{e}^i\right)}{1-\delta(1-\lambda)}+\frac{\delta\lambda}{1-\delta(1-\lambda)}w^i\left(\bar{e}^i\right). \tag{37}$$

Now substitute (6) with equality for $v^i\left(\bar{e}^i\right)$ in (37). Rearranging terms, we obtain:

$$\frac{\delta(1-\delta)(1-\lambda)}{1-\delta(1-\lambda)}w^i\left(\bar{e}^i\right)=\frac{\pi\left(\bar{e}^i,\bar{e}^i\right)}{1-\delta(1-\lambda)}-\pi\left(0,\bar{e}^i\right).$$

Since $\bar{e}^i$ is decreasing with p^i, $w^i\left(\bar{e}^i\right)$ increases with p^i if and only if,

$$\frac{\partial d\left(\bar{e}^i,0\right)}{\partial\bar{e}^i}\leq 0.$$

Proof of Proposition 3

We now show that the efficient segregating PPE, as defined in Definition 1, is indeed a PPE in pure strategies, and that vector (p^i) exists for all matrices of symbol switching costs, $\left(c\left(i,j\right)\right)_{i,j}$, such that the efficient segregating PPE can be sustained as an equilibrium. First, by Proposition 1, the equilibrium efforts in (9) are always well defined under condition 1. Given that all deviations are met with a partnership termination, no unilateral deviation from $\bar{e}^i$ can increase a player's continuation value by construction. And because all deviations in termination decisions after a deviation are met with partnership termination, this equilibrium is also subgame perfect. Finally, we need that the condition in Proposition 3 that excludes symbol switching,

$$w^j\left(\bar{e}^j\right)-w^i\left(\bar{e}^i\right)\leq\frac{c\left(i,j\right)}{\delta},$$

is satisfied for some vector (p^i) for all matrices of symbol switching costs, $\left(c\left(i,j\right)\right)_{i,j}$. Note that for a vector of equal components, $p^i=p^j=p$ for all i,j, we have $w^i\left(\bar{e}^i\right)=w^j\left(\bar{e}^j\right)$, which implies that the inequality in Proposition 3 is satisfied for all $\left(c\left(i,j\right)\right)_{i,j}$.

Derivation of (13) and (14)

We briefly illustrate the derivation of (13). Note that after k rounds of equilibrium play, the expected continuation value on the equilibrium path is

$$\begin{aligned}&\pi\left(e_k,e_k\right)+\delta\left(\left(1-\lambda\right)\pi\left(e_{k+1},e_{k+1}\right)+\lambda\pi\left(e_0,e_0\right)\right)\\&+\delta^2\left(\left(1-\lambda\right)^2\pi\left(e_{k+2},e_{k+2}\right)+\lambda\left(1-\lambda\right)\pi\left(e_1,e_1\right)+\lambda^2\pi\left(e_0,e_0\right)\right)\\&+\delta^3\left(\left(1-\lambda\right)^3\pi\left(e_{k+3},e_{k+3}\right)+\lambda\left(1-\lambda\right)^2\pi\left(e_2,e_2\right)+\lambda^2\left(1-\lambda\right)\pi\left(e_1,e_1\right)+\lambda^3\pi\left(e_0,e_0\right)\right)\\&+...\end{aligned}\tag{38}$$

The expected continuation value of cheating in the k-th is

$$\begin{aligned}&\pi\left(0,e_k\right)+\delta\pi\left(e_0,e_0\right)+\delta^2\left(\left(1-\lambda\right)\pi\left(e_1,e_1\right)+\lambda\pi\left(e_0,e_0\right)\right)\\&+\delta^3\left(\left(1-\lambda\right)^2\pi\left(e_2,e_2\right)+\left(1-\lambda\right)\lambda\pi\left(e_1,e_1\right)+\lambda^2\pi\left(e_0,e_0\right)\right)\\&+...\end{aligned}\tag{39}$$

Incentive compatibility requires that the difference between (38) and (39) is positive. After some algebraic manipulation, we obtain the constraint in (13).

For the objective function in (14), note that at the 0-th round of cooperation, a player wishes to maximize

$$\begin{aligned}
&\pi(e_0,e_0)+\delta(1-\lambda)\pi(e_1,e_1)+\delta\lambda\pi(e_0,e_0)\\
&+\delta^2(1-\lambda)^2\pi(e_2,e_2)+\delta^2\left(\lambda^2+(1-\lambda)\lambda\right)\pi(e_0,e_0)+\delta^2(1-\lambda)\lambda\pi(e_1,e_1)\\
&+...\\
&=\pi(e_0,e_0)\left(1+\lambda\sum_{t=1}^{\infty}\delta^t\right)+(1-\lambda)\delta\pi(e_1,e_1)\left(1+\lambda\sum_{t=1}^{\infty}\delta^t\right)\\
&+((1-\lambda)\delta)^2\pi(e_2,e_2)\left(1+\lambda\sum_{t=1}^{\infty}\delta^t\right)+...\\
&=\sum_{k=0}^{\infty}\pi(e_k,e_k)(\delta(1-\lambda))^k\left(1+\lambda\sum_{t=1}^{\infty}\delta^t\right)\\
&=\left(1+\frac{\lambda\delta}{1-\delta}\right)\sum_{k=0}^{\infty}\pi(e_k,e_k)(\delta(1-\lambda))^k.
\end{aligned}$$

That means that we seek to maximize $\sum_{k=0}^{\infty}(\delta(1-\lambda))^k\pi(e_k,e_k)$

Proof of Proposition 4

The proof follows the lines of the proof of Proposition 1. In particular, we can use (20) and using efficiency players will select the largest e such that the inequality is satisfied with equality. This yields the expression in (21). Furthermore, after rearranging (20) (and after we impose equality), we obtain:

$$\frac{\pi\left(e_1^{GTB},e_1^{GTB}\right)}{\pi\left(0,e_1^{GTB}\right)}=\frac{1-\delta(1-\lambda)}{1-(\delta(1-\lambda))^{T+1}}.$$

And using similar arguments as in the proof for Proposition 1, the equilibrium efforts in (21) are always well defined under condition 1. Moreover, since $\frac{1-\delta(1-\lambda)}{1-(\delta(1-\lambda))^{T+1}}$ is decreasing with T, given the properties of the ratio $\frac{\pi(e,e)}{\pi(0,e)}$ and the max operator defining equilibrium efforts, we obtain that e_1^{GTB} is increasing with T. We can also again use the same reasoning as in the proof of Proposition 1 to obtain that equilibrium efforts are increasing and right-continuous with respect to δ and decreasing and left-continuous with respect to λ.

Proof of Proposition 5

Consider a particular T-GTB, i.e. a gradual trust-building PPE with an incubation phase length given by T. First, note that the equilibrium efforts associated with the T-GTB equilibrium are given by (21). Now, notice that the same effort levels would arise in an efficiently segregating PPE with a uniform symbols' distribution, $\alpha = (1/n, ..., 1/n)$, if

$$n = \bar{n} = \frac{1 - (\delta(1-\lambda))^{T+1}}{(\delta(1-\lambda))^{T}[1 - \delta(1-\lambda)]}. \tag{40}$$

Notice that we ignore the integer restriction $\bar{n} \in \mathbb{N}_0$ for now. Substituting $\alpha^i = 1/n$ for all s^i and imposing stationarity, $x^i_{t+1} = x^i_t$ in (2) yields the following:

$$x^i = x = \frac{\lambda}{n\lambda + 1 - \lambda}. \tag{41}$$

Substituting (41) into the expression for the average expected payoffs, (24), we obtain:

$$W^i((1/n, ..., 1/n); n) = \frac{\lambda}{n\lambda + 1 - \lambda}\frac{1}{n}\frac{\pi(0, \bar{e})}{1-\delta} + \frac{1-\lambda}{n(n\lambda + 1 - \lambda)}\left[\frac{\pi(\bar{e}, \bar{e})}{1 - \delta(1-\lambda)} + \frac{\delta\lambda}{1-\delta(1-\lambda)}\frac{1}{n}\frac{\pi(0, \bar{e})}{1-\delta}\right], \tag{42}$$

We are now interested in the expression for $W^i((1/n, ..., 1/n); n) - W_{GTB}(T; \delta, \lambda)$ for all s^i. First, we substitute (40) into (42) such that the difference between average expected payoffs in the associated efficiently segregating PPE and the T-GTB PPE is only a function of T, δ and λ. In particular, after some algebra, it can be shown that $W^i((1/n, ..., 1/n); n) - W_{GTB}(T; \delta, \lambda) \geq 0$ if and only if $F(T; \delta, \lambda) \geq 0$, with:

$$\begin{aligned} F(T; \delta, \lambda) = \; & (1-\lambda)(1-\delta)\left[1 - (\delta(1-\lambda))^{T+1}\right]\left[1 - \delta(1-\lambda)\frac{1-(\delta(1-\lambda))^{T}}{1-(\delta(1-\lambda))^{T+1}}\right] \\ & + \lambda\left[1 - (\delta(1-\lambda))^{T+1}\right] - (\delta(1-\lambda))^{T}[1 - \delta(1-\lambda)](1-\delta). \end{aligned} \tag{43}$$

Notice that $\lim_{\epsilon \to 0^+} F(T; \epsilon, \lambda) = 1$ and $F(T; 1, \lambda) = \lambda\left(1 - (1-\lambda)^{T+1}\right)$, for all $T > 0$ and $\lambda \in (0, 1)$. Furthermore, $\frac{\partial F(T;\delta,\lambda)}{\partial \delta} < 0$ for all $T > 0$ and $\lambda \in (0, 1)$ and $\frac{\partial F(T;\delta,\lambda)}{\partial \lambda} > 0$ for all $T > 0$ and $\delta \in (0, 1)$, which shows that $W^i((1/n, ..., 1/n); n) - W_{GTB}(T; \delta, \lambda)$ is decreasing in δ and increasing in λ.

Proof of Proposition 6

In the case of a best-GTB, the length of the incubation phase is given by T^* as defined in (23). For given values of $\delta, \lambda \in (0, 1)$, we can again find a uniform symbols' distribution $\alpha = (1/n, ..., 1/n)$ such that

$$n = \bar{n} = \frac{1 - (\delta(1-\lambda))^{T^*+1}}{(\delta(1-\lambda))^{T^*}[1 - \delta(1-\lambda)]}. \tag{44}$$

The rest of the proof is very similar to the proof of Proposition 5, in that for the given levels of δ, λ we again obtain the expression (43). Notice that, given that T^* is now a function of δ, λ, the comparative statics are a bit more involved. In particular, given that for a fixed value for λ, say $\lambda = \bar{\lambda}$, the map $T^*\left(., \bar{\lambda}\right)$ is a step function. Hence, the difference $W^i\left(\left(1/n, ..., 1/n\right); n\right) - W_{GTB}\left(T; \delta, \lambda\right)$ is continuously decreasing, except for jumps downward at levels δ for which T^* changes.

Proof of Proposition 8

Suppose that players agree on a non-constant effort plan $(e_\tau)_{\tau=0,1,...}$. This equilibrium sequence constitutes the outside option at all moments of the current partnership (i.e., what to expect in the next partnership) and is thus independent of how far a player is in his current partnership. If at some τ a certain effort e is sustainable, then all efforts in time periods τ for which $e_\tau \leq e$ can be renegotiated to level e. Repeat this argument and conclude that the only effort that is robust to $\varepsilon-$renegotiation is a constant and efficient effort level, where efficiency means exhausting the incentive compatibility constraints. Note that this argument holds for any $\varepsilon > 0$. Now, denote such constant efficient effort level by $\tilde{e}$. Let $v\left(\tilde{e}\right)$ denote the expected continuation value. Then,

$$v\left(\tilde{e}\right) = \pi\left(\tilde{e}, \tilde{e}\right) + \delta v\left(\tilde{e}\right).$$

In order for $\tilde{e}$ to be incentive compatible and efficient, we should have that:

$$v\left(\tilde{e}\right) = \pi\left(0, \tilde{e}\right) + \delta v\left(\tilde{e}\right).$$

But this means that $\pi\left(0, \tilde{e}\right) = \pi\left(\tilde{e}, \tilde{e}\right)$, which can only hold for $\tilde{e} = 0$. This is the (trivial) PPE associated with a an equilibrium strategy, σ, such that all players constantly exert no efforts in any partnership. Now, consider a small $\varepsilon > 0$ and let $\hat{e}\left(\varepsilon\right)$ be a (positive) effort level that is part of a jointly deviating strategy σ' such that the distance $M\left(\sigma, \sigma'\right) = \varepsilon$. A joint deviation from zero efforts to $\hat{e}\left(\varepsilon\right)$ gives the following expected continuation payoff:

$$v\left(\hat{e}\left(\varepsilon\right)\right) = \frac{\pi\left(\hat{e}\left(\varepsilon\right), \hat{e}\left(\varepsilon\right)\right)}{1 - \delta(1-\lambda)},$$

and incentive compatibility requires that

$$v\left(\hat{e}\left(\varepsilon\right)\right) \geq \pi\left(0, \hat{e}\left(\varepsilon\right)\right),$$

which is clearly viable. Hence, we can only sustain $\tilde{e} = 0$ for $M(\sigma, \sigma') \leq \varepsilon$ in the non-generic case where $\delta(1-\lambda) = 0$. Generically, there does not exist a symbol-blind PPE which is robust to a a joint deviation of (arbitrarily small) size ε.

Proof of Proposition 9

Let σ be an equilibrium strategy for a player as part of an efficiently segregating PPE. We will put bounds on $\varepsilon > 0$ such that, the efficiently segregating PPE is robust against joint deviations prescribing for each player a strategy σ' and $M(\sigma, \sigma') \leq \varepsilon$, i.e. for all $\sigma' \in \mathcal{N}(\sigma; \varepsilon)$. First, notice that, in homogeneous s^i partnerships, two players who jointly deviate to σ', with $M(\sigma, \sigma') \leq \varepsilon$, with an associated effort level e, obtain a continuation value

$$v^h(e) = \pi(e, e) + \delta(1-\lambda)v^h(e) + \delta\lambda w^i\left(\bar{e}^i\right). \tag{45}$$

Solving (45) for $v^h(e)$ and substituting into the incentive compatibility constraint

$$v^h(e) \geq \pi(0, e) + \delta w^i\left(\bar{e}^i\right),$$

we obtain that a joint deviation to e is only viable if

$$d(e, 0) \geq \delta \frac{(1-\delta)(1-\lambda)}{1-\delta(1-\lambda)} w^i\left(\bar{e}^i\right). \tag{46}$$

Note then that (46) is satisfied with equality for $\bar{e}^i$ by construction, and that $d(e, 0)$ decreases with e for all $e \geq e^*$, such that if $p^i \leq p^*$ for all i, it cannot be that (46) is satisfied for any $e \geq \bar{e}^i$. Hence, a joint deviation to a higher effort level in a homogeneous partnership is never viable if $p^i \leq p^*$ for all i.

Second, consider a pair of players with different symbols, say an s^i and s^j player. Suppose without loss of generality that $p^i \leq p^j < p^*$. From Lemma 2, it follows that $w^i(\bar{e}^i) \leq w^j(\bar{e}^j)$. We now show that the assumption (33) is sufficient to ensure that players don't want to deviate jointly to $\hat{e}(\varepsilon)$. To that end, note that the continuation value in a joint heterogeneous deviation with effort level $\hat{e}(\varepsilon)$ is given by:

$$v^H(\hat{e}(\varepsilon)) = \pi(\hat{e}(\varepsilon), \hat{e}(\varepsilon)) + \delta(1-\lambda)v^H(\hat{e}(\varepsilon)) + \delta\lambda w^i\left(\bar{e}^i\right),$$

from which we obtain that joint deviation to heterogeneous cooperation is viable if

$$\frac{\pi(\hat{e}(\varepsilon), \hat{e}(\varepsilon))}{1-\delta(1-\lambda)} - \pi(0, \hat{e}(\varepsilon)) \geq \frac{\delta(1-\lambda)(1-\delta)}{1-\delta(1-\lambda)} w^i\left(\bar{e}^i\right). \tag{47}$$

As a result of Lemma 2, the outside continuation value $w\left(\bar{e}^i\right)$ is increasing with p^i; therefore, if we use $\hat{p}\left(\varepsilon\right)$ to denote the infimum of all $p^i \leq p^*$ satisfying (33) (i.e., not satisfying (47), we see that our efficiently segregating PPE is sustainable if $p^i > \hat{p}\left(\varepsilon\right)$. Moreover, $\hat{p}\left(\varepsilon\right)$ is increasing in ε. By increasing ε, the upper bound of joint deviations, $\hat{e}\left(\varepsilon\right)$ increases. An efficiently segregating PPE is sustained as long as $\hat{e}\left(\varepsilon\right) < \bar{e}^i \leq e^*$ for all i. When $\hat{e}\left(\varepsilon\right) = e^*$, then $\hat{p}\left(\varepsilon\right) = p^*$, and therefore, the interval of feasible shares p^i becomes empty and the efficient segregating PPE is no longer robust to joint deviations of size ε.