MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Social learning under heterogeneous preferences“

verfasst von / submitted by

Sebastian Ertner, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2022 / Vienna 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 913
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Volkswirtschaftslehre
Betreut von / Supervisor:	Univ.-Prof. Dipl.-Math. Karl Schlag, PhD

Abstract

We study a model of social learning with two types of boundedly rational individuals who differ in preferences and who repeatedly choose between two risky alternatives according to simple behavioral rules. We prove that no rule exists that always leads both types to choosing their respective preferred alternative. Such convergence can only be guaranteed if both types prefer the same alternative. Among all rules that lead to optimal behavior in the setting where there is only one type, we identify those that minimize maximum regret in the short run in the setting with two types. These rules prescribe to either imitate another individual that achieved a better result or to change actions more often the more dissatisfied an individual is with their own result.

Zusammenfassung

Wir untersuchen ein Modell des sozialen Lernens mit zwei Typen von beschränkt rationalen Individuen, die sich in ihren Präferenzen unterscheiden und die wiederholt gemäß einfachen Entscheidungsregeln zwischen zwei riskanten Alternativen wählen. Wir beweisen, dass keine Regel existiert, die immer dazu führt, dass beide Typen die jeweils bevorzugte Alternative wählen. Solch eine Konvergenz kann nur garantiert werden, wenn beide Typen dieselbe Alternative bevorzugen. Unter allen Regeln, die zu optimalem Verhalten im Modell mit nur einem Typen führen, identifizieren wir jene, die im Modell mit zwei Typen auf kurze Sicht den maximalen Schaden minimieren. Jene Regeln schreiben vor, dass entweder ein anderes Individuum imitiert werden sollte, welches ein besseres Ergebnis erzielt hat, oder dass Individuen desto öfter ihre Wahl wechseln sollten, je unzufriedener sie mit ihrem Ergebnis sind.

1 Introduction

In this master thesis we study a model of social learning under heterogeneous preferences.[1] We closely follow the model presented in Schlag (1998) and introduce the following extension. In said publication the population is assumed to be homogeneous. Here we instead assume there exist multiple types that differ in their expected valuation of the available alternatives.

We motivate our model with a simple real world application from Schlag (1998) where individuals repeatedly face the decision problem of choosing which one of two restaurants to visit. In the context of our model, such a visit has a stochastic outcome - multiple factors determine the utility gained from it, not only the inherent quality of the restaurant itself. In this sense, a particular visit of the preferred restaurant can randomly turn out worse than a visit of the less preferred one. We could imagine that by chance the preferred restaurant on a particular day is not able to serve the individual's favorite dish due to supply issues or that the cook accidentally spoils the meal. That being said, individuals will in general still prefer one restaurant over the other in terms of expected utility.
In Schlag (1998) the population is homogeneous - everyone prefers the same restaurant. This will be the case here only to a certain extent. We allow for differing preferences, which we introduce as types for which the distribution over outcomes for each alternative is different. Preferences will in general still be somewhat similar and not too far from the homogeneous case. In the terms of our example we can imagine that either type will still place significant value on the quality of food regardless of cuisine or other characteristics of the restaurants. Thus heterogeneity in preferences will in practice be limited. In our analysis we nevertheless allow for strongly differing tastes, including the extreme case that the types only value distinct restaurants.
For the reason of simplicity we assume that there are only two types of individuals in the population. Then the two types will differ in their valuations for the two restaurants, which correspond to the expected utility of a visit to either of them. Even if both individuals prefer the same restaurant, their expected valuations in general do not need to be the same.
We assume the individuals to be boundedly rational. They do not form be-

[1] I would like to thank my supervisor Prof. Karl Schlag for many helpful discussions and for his guidance and support during the writing of this master thesis.

liefs about the outcome distribution for either alternative. Further, they only remember their last action as well as the realized outcome and are informed of the last action as well as the realized outcome of one random other individual. In the context of our model this would mean that individuals only remember which restaurant they visited last and the outcome of that visit. Further they get information from someone else of unknown type about their last restaurant visit and outcome. As we assume that this other individual is chosen randomly from the population, this assumption most closely resembles the case of consulting anonymous online reviews of restaurants.
Individuals decide which of the two restaurants to visit next based on this information. We assume that individuals use a 'behavioral rule' to make this decision. Such a rule prescribes an action according to a simple ad-hoc procedure, given the information available as described above. The rule can prescribe a mixed action, in which case the individuals will select the next alternative randomly, given the probabilities prescribed by the behavioral rule. We further assume that all individuals of both types use the same behavioral rule. We limit our attention to behavioral rules that are imitating, which means that individuals only consider playing a different alternative if they observe another individual doing so.

In our analysis of this model we will first study the properties of the population dynamics. We will see that contrary to the setting in Schlag (1998), there do not exist *improving* rules that always lead the population to the state in which all individuals choose their preferred alternative with certainty. However, we will show that this is still the case if both types prefer the same alternative. We will then assess the performance of the optimal rules from the homogeneous setting in this heterogeneous generalization. We evaluate them based on a short run measure of minimax regret from one round to the next. Similar to Schlag (1998), we find that the *Proportional Imitation Rule* will be optimal in this setting. This rule prescribes to imitate the observed action proportional to the difference in payoffs if the other action did yield a higher outcome. However, contrary to there, the *Proportional Reviewing Rule* will perform equally well. This rule does not take the payoff of the observed other individual into account when prescribing the action for the next round. Instead, individuals under this rule change their action more frequently if they are dissatisfied with their own result.

We now briefly present related literature. Kirman (2006) emphasizes the importance of heterogeneity in individuals' characteristics for economic activity, and argues that this fact should be reflected in economic models. Köster et al. (2020) model heterogeneity in preferences in a way where agents have an individual preference for one particular variety of a good, but also gain utility (albeit less so) from consuming other varieties. Björnerstedt and Schlag (1996) study a similar two-armed bandit model where different behavioral rules compete against each other. Schlag (2004) similarly has individuals choose rules that minimize maximum regret in the same setting as in the original publication Schlag (1998).

The document is organized as follows. In section 2 we introduce the model setting. In section 3 we present relevant results that have been established in Schlag (1998) and apply them to our setting. In section 4 we derive the population dynamics for our model setting. In section 5 we prove that improving rules cannot exist in our setting, as opposed to the homogeneous case. In section 6 we study stationarity in our model and derive conditions on the parameter configuration for the existence of an internal rest point. In section 7 we determine the conditions for convergence of the population towards either both types' preferred alternative or to another rest point. Further we prove that the population will always converge towards an alternative if it is preferred by both types. In section 8 we evaluate the performance of different decision rules in our model in the short-run using a regret measure and conclude that PIR and PRR are the optimal rules. Section 9 concludes this document by summarizing our findings and discussing possible directions for further research.

2 Model setting

The model is a generalization of the model presented in Schlag (1998). Our assumptions closely follow the setting as presented there.

In our model, players face an infinitely repeated decision problem in the form of a two-armed bandit. In other words, players can choose in every round one out of two independent alternatives $C \in A = \{T, B\}$ with unknown payoff distributions. The population is assumed to be infinite and consists of two

types of players[2] $i \in I = \{1, 2\}$ with respective population shares $\lambda_i \in [0, 1]$ with $\sum_i \lambda_i = 1$. Note that for a population share of $\lambda_i = 1$ for one type $i \in I$ the model coincides with the one presented in Schlag (1998), which we will throughout refer to as the *homogeneous case* or the *homogeneous setting*.
The two types differ in their payoff distributions $(P^i_C)_{C\in A}$ over a known finite support in $[\alpha, \omega]$ that is common among the alternatives $C \in A$ and among both types. Players have limited memory - the only information available to them comes from the last round of play and includes the following: their own realized action $C \in A$, their own realized payoff $x \in [\alpha, \omega]$ as well as the action $D \in A$ and realized payoff $y \in [\alpha, \omega]$ of a random other player of unknown type. Players use this information to determine their action in the current round according to a behavioral rule[3] $F(C, x, D, y)$ that prescribes a pure or mixed action. Importantly, we assume that all individuals use the same rule. Note that due to players only remembering their realized action, there is no possibility for information transfer from one round to the next on an individual level. The learning occurs on the population level.
We define for the expected payoffs of type i:

$$\chi_i := \mathbb{E}[u_i(T)] = \textstyle\sum_x x P^i_T(x) \quad \text{and} \quad \mu_i := \mathbb{E}[u_i(B)] = \textstyle\sum_x x P^i_B(x)$$

With these definitions it follows that $T \succ_i B \Leftrightarrow \chi_i > \mu_i$. For ease of use we normalize the minimum α and maximum ω in the support of the payoff distributions to 0 and 1 using a positive affine transformation. Therefore it will be the case in the remainder of this document that $\chi_i, \mu_i \in [0, 1]$ for both types. See Fig. 1 for an example of a decision problem of the specified form. Note how the alternatives are independent and how payoffs are normalized. Further it is the case in this particular example that the probability of receiving the better payoff is equal to the expected payoff of that alternative.

To conclude this section we summarize that what we have presented here is the *heterogeneous setting*. It is an extension of the homogeneous one to multiple types that differ in their payoff distributions. The homogeneous setting results from it as an edge case if either $\lambda_i = 1$ for one $i \in I$ or if $\mathbb{E}[u_i(C)]$ is equal for both types for all $C \in A$.

[2] A note on notation: As the model is symmetric in both types, we use i within our arguments to identify one arbitrary type and refer to the other type by j. When needed we then further use k and l to refer to a player that may be of either type.

[3] See section 3.1 for the definition.

	$\mu_i \chi_i$	$(1-\mu_i)\chi_i$	$\mu_i(1-\chi_i)$	$(1-\mu_i)(1-\chi_i)$
T	1	1	0	0
B	1	0	1	0

Figure 1: Example of a payoff distribution of a decision problem with two independent risky alternatives for a player of type i. The top row of each column represents the probability of the payoff of that column being realized.

3 Concepts

In this section we recall results from the homogeneous case presented in Schlag (1998) and adjust them to the heterogeneous setting. Within this section, only these adjustments are novel.

3.1 Behavioral rules

In the following we use the terms *behavioral rule* and *decision rule* interchangeably. A behavioral rule F is a function that assigns to every quadruple of own realized action $C \in A$ with realized payoff $x \in [\alpha, \omega]$ and observed action $D \in A$ with realized payoff $y \in [\alpha, \omega]$ a mixed action in A:

$$F(C, x, D, y) : A \times [\alpha, \omega] \times A \times [\alpha, \omega] \to \Delta(A)$$

We denote by $F_E(C, x, D, y) \in [0, 1]$ the probability of playing action $E \in A$ in the next round.
In our analysis we limit our search to behavioral rules that are *imitating*, a choice that we will justify in section 3.3. An imitating or *imitation rule* is a rule with $F_E(C, x, D, y) = 0$ for all $E \notin \{C, D\}$ and all $x, y \in [\alpha, \omega]$. Given that there are only two alternatives in our model, this means that the only non-zero switching probabilities will be $F_T(B, x, T, y)$ and $F_B(T, x, B, y)$.

$F_T^{ik}(B, T)$ is then the expected average probability that an individual of type i that played B and observed an individual of type $k \in \{i, j\}$ that played T will switch to T, i.e.:

$$F_T^{ik}(B, T) = \sum_x \sum_y P_B^i(x) P_T^k(y) F_T(B, x, T, y) \quad (1)$$

Similarly for $F_B^{ik}(T, B)$. As we will refer to these expected switching probabilities frequently, we introduce a shorthand notation that we will use in the remainder of this document:

$$s_T^{ik} := F_T^{ik}(B, T) \text{ and } s_B^{ik} := F_B^{ik}(T, B)$$

For clarity we further define $s_T^i := s_T^{ii}$ and $s_B^i := s_B^{ii}$. Note that these last switching probabilities coincide with the homogeneous case.

3.2 Population dynamics in the homogeneous case

We characterize the population dynamics in terms of (expected) player proportions. By $z_T^{i,t} \in [0, 1]$ we identify the share of players of type i who play alternative T in round t. The share of players of type i who play alternative B is given by $z_B^{i,t} = 1 - z_T^{i,t}$. Note that $z_T^{i,t}$ is the realization of a random variable $Z_T^{i,t}$, as matching and selection of the action for the next round are random events. To study the evolution of the whole population, one state variable for each type is sufficient - denoted in the following by $z_T^t = (z_T^{i,t}, z_T^{j,t})$. In particular, we do not need to consider the realized action of each individual in a given round, as we have assumed an infinite population. Note that $z_T^t \in [0, 1]^2$. We refer to the point $z_T^t = (0, 0)$ as the B *corner* and to $z_T^t = (1, 1)$ as the T *corner*. Both points are labeled the *symmetric corners*. By contrast, we refer to the points $z_T^t = (0, 1)$ and $z_T^t = (1, 0)$ as the *asymmetric corners*.
The population dynamics give us the expected value of the share Z_T^{t+1} in the next round as a function of the current share z_T^t as well as the expected switching probabilities implied by the used imitation rule and the parameter configuration. In the homogeneous edge case of our model setting the population dynamics are:[4]

$$\mathbb{E}[Z_T^{i,t+1}] = z_T^{i,t} + z_T^{i,t}(1 - z_T^{i,t})[s_T^i - s_B^i] \tag{2}$$

As we assume an infinite population, the actual realized value will coincide with the expected value:

$$z_T^{i,t+1} = \mathbb{E}[Z_T^{i,t+1}]$$

[4]See appendix A.1 for a derivation of the population dynamics in the heterogeneous case, which is a generalization of the homogeneous case for which $\lambda_i = 1$.

3.3 Improving rules in the homogeneous case

We are interested in optimal behavior and thus we want to find rules that improve on the status quo.
In Schlag (1998) the term *improving* is chosen for rules that for any decision problem and in any state yield a weakly higher expected payoff than the rule called Never Switch. We adopt this definition and call rules that have this property in the homogeneous setting *homogeneously improving rules*. Equivalently for the heterogeneous setting. Note that the property of being heterogeneously improving implies that of being homogeneously improving. We will prove in section 5 that heterogeneously improving rules do not exist. Never Switch, as the name suggests, designates the behavioral rule of never switching away from the played alternative, i.e. $F_D(C, x, D, y) = 0$ for all $C, D \in A$ and all $x, y \in [\alpha, \omega]$. Following Theorem 1 in Schlag (1998), a rule is improving in the homogeneous setting if and only if it has the properties of being *imitating* and if for all χ_i, μ_i there exists a so-called *switching rate* $\sigma \in [0, 1]$ such that:

$$s_T^i - s_B^i = \sigma(\chi_i - \mu_i) \tag{3}$$

Due to this result we will henceforth only consider imitation rules. The reason for requiring (3) can be intuitively understood by considering the population dynamics (2). If $s_T^i > s_B^i$, then the population will converge towards $z_T^{i,t} = 1$, which is only desirable if T gives higher expected payoff, i.e. if $\chi_i > \mu_i$.

Per definition, Never Switch is an improving rule. However, it never improves strictly on the current situation as $\sigma = 0$ for it. Thus it is termed a *degenerate* improving rule.
We are only interested in non-degenerate or *strictly improving rules*, which conversely are rules with $\sigma > 0$. We identify the rules for which this holds in the following.

To simplify the analysis we focus on imitation rules for extreme payoffs $x, y \in \{0, 1\}$ for the remainder of this document. Thus there are only four distinct pairs of own and observed outcome. This implies that an imitation rule for such extreme payoffs can be fully identified by four parameters.

For the sake of brevity we introduce a shorthand notation:

$$\begin{aligned} a &:= F_T(B,1,T,1) = F_B(T,1,B,1) \\ b &:= F_T(B,0,T,1) = F_B(T,0,B,1) \\ c &:= F_T(B,1,T,0) = F_B(T,1,B,0) \\ d &:= F_T(B,0,T,0) = F_B(T,0,B,0) \end{aligned}$$

We do not distinguish between switching from T to B and switching from B to T as we assume that players have no ex-ante information about the alternatives that would allow them to rationally discriminate among them.

Note that even though we limit our analysis to such rules for extreme payoffs, we can still prescribe behavior for interior payoffs $x, y \in (0,1)$ according to the same rules. This can be done by employing the "randomization trick" presented in Schlag et al. (2013). For this trick we add another step to each player's procedure of determining the action for the next round. Each player transforms the observed payoffs $x, y \in [0,1]$ into extreme payoffs $\hat{x}, \hat{y} \in \{0,1\}$ by performing random draws such that $\hat{x}$ is equal to 1 with probability x and equal to 0 with probability $1-x$ and similarly for $\hat{y}$.[5] They then treat $\hat{x}, \hat{y}$ as if those were the payoffs they observed and apply their imitation rule for extreme payoffs. Such a rule will thus have the following form:

$$\begin{aligned} F_T(B,x,T,y) = {} & xyF_T(B,1,T,1) + x(1-y)F_T(B,1,T,0) + \\ & + (1-x)yF_T(B,0,T,1) + (1-x)(1-y)F_T(B,0,T,0) = \\ & = xya + (1-x)yb + x(1-y)c + (1-x)(1-y)d \end{aligned}$$

Similar for $F_B(T,x,B,y)$. Given this simplification, the expected switching probabilities according to (1) are:

$$\begin{aligned} s_T^{ik} &= F_T^{ik}(B,T) = \mu_i\chi_k a + (1-\mu_i)\chi_k b + \mu_i(1-\chi_k)c + (1-\mu_i)(1-\chi_k)d \\ s_B^{ik} &= F_B^{ik}(T,B) = \chi_i\mu_k a + (1-\chi_i)\mu_k b + \chi_i(1-\mu_k)c + (1-\chi_i)(1-\mu_k)d \end{aligned} \tag{4}$$

[5] Recall that we normalize the support of the payoffs x, y from $[\alpha, \omega]$ to $[0,1]$ via a positive affine transformation.

For the homogeneous case we find:

$$\begin{aligned}
s^i_T - s^i_B &= \mu_i\chi_i a + (1-\mu_i)\chi_i b + \mu_i(1-\chi_i)c + (1-\mu_i)(1-\chi_i)d - \\
&\quad - [\mu_i\chi_i a + \mu_i(1-\chi_i)b + (1-\mu_i)\chi_i c + (1-\mu_i)(1-\chi_i)d] = \\
&= \mu_i(1-\chi_i)c + (1-\mu_i)\chi_i b - \mu_i(1-\chi_i)b - (1-\mu_i)\chi_i c = \\
&= (\chi_i - \mu_i)(b-c)
\end{aligned}$$

Comparison to (3) implies that an imitation rule is improving in the homogeneous edge case if it holds that the switching rate σ is equal to $b - c > 0$, but there are no conditions on a and d. Note that any rule that shall be improving in the general heterogeneous setting also has to fulfill this condition as the homogeneous case is a special case of it.

3.4 Dominant rules in the homogeneous case

It is shown in Schlag (1998) that a higher switching rate results in better performance of an improving rule. Thus the so-called *dominant rules* are there defined as those improving rules that always yield a weakly higher expected payoff than all other improving rules. That means dominant rules have the maximal possible switching rate $\sigma = 1$, which corresponds to $b = 1$ and $c = 0$. Given that there are no conditions on a and d, we can identify four fundamental dominant rules. All other dominant rules can be expressed as convex combinations of these four. See Table 1 for a characterization.[6] The rules are introduced and named in Schlag (1998), Schlag (1999) and Schlag (2022).

The *Proportional Imitation Rule* (PIR) is the unique dominant rule that never imitates an action that attained a lower payoff, as Theorem 2 in Schlag (1998) states. Thus this rule never switches unless the other action received a higher payoff, in which case it switches with probability equal to the difference in payoffs.

The *Proportional Observation Rule* (POR) is the unique dominant rule that does not require observation of the own payoff (but of one's own action). It always switches with probability equal to the other player's realized payoff.

The *Proportional Reviewing Rule* (PRR) is the unique dominant rule that

[6] Note that we here list the randomization trick equivalents for the rules PIR and RPIR, as signified by the leading "$\rho-$". POR and PRR have the same representation under the randomization trick.

does not require observation of the payoff of another player (but of their action). It always switches with probability equal to one minus the own payoff. The fourth fundamental dominant rule is the *Reverse Proportional Imitation Rule* (RPIR). It always switches, unless the own realized payoff is higher than the observed payoff, in which case the switching probability is equal to one minus that difference in payoffs.

Rule	(a,d)	$F_D(C, x, D, y)$	s_T^{ik}	s_B^{ik}
ρ - PIR	(0,0)	$(1-x)y$	$(1-\mu_i)\chi_k$	$(1-\chi_i)\mu_k$
POR	(1,0)	y	χ_k	μ_k
PRR	(0,1)	$1-x$	$1-\mu_i$	$1-\chi_i$
ρ - RPIR	(1,1)	$1-x(1-y)$	$1-\mu_i(1-\chi_k)$	$1-\chi_i(1-\mu_k)$

Table 1: Overview of the four fundamental dominant rules of the homogeneous setting. Notice that all four rules have the property $s_T^i - s_B^i = \sigma(\chi_i - \mu_i)$ with $\sigma = 1$. Further note that for PRR $s_C^{ik} = s_C^i$.

We have summarized in this section the results from Schlag (1998) that we intend to study in the heterogeneous setting. In the setting there, which we have been referring to as the homogeneous setting, there exists the class of behavioral rules referred to as improving rules, which give weakly higher expected payoff than the rule Never Switch. Further, among these improving rules there exist a subclass of so-called dominant rules, that generate weakly higher expected improvement than any other improving rule.
We will now follow the same analysis in the heterogeneous setting. First we derive the population dynamics. In the next step we will see that improving rules cannot exist in this setting. This forces us to follow a different approach than Schlag (1998) in regard to rule selection. We will evaluate the fundamental dominant rules from the homogeneous setting in the heterogeneous one. We justify this approach as we are considering social learning through imitation. It only makes sense for imitation to occur if individuals can learn something valuable from each other. That means that preferences will in general be similar. Although we assume that heterogeneity is present, it will be limited and not far from the homogeneous case. Thus we evaluate how

homogeneous rules fair in a generalized setting.
To conclude this section we emphasize once more that we only consider imitation rules for extreme payoffs in the remainder of this document.

4 Population dynamics

We derive the population dynamics for our model (see appendix A.1 for details). There are two equivalent representations, the first of which is:

$$\begin{aligned} z_T^{i,t+1} - z_T^{i,t} = \lambda_i z_T^{i,t}(1 - z_T^{i,t})[s_T^i - s_B^i] + \\ + (1 - \lambda_i)[(1 - z_T^{i,t}) z_T^{j,t} s_T^{ij} - z_T^{i,t}(1 - z_T^{j,t}) s_B^{ij}] \end{aligned} \tag{5}$$

This representation has the following interpretation: The first term represents the change in the share of T players in type population i that is induced by observing another player of the own type, while the second term represents the same when observing a player of the other type.

The alternative representation of the population dynamics is:

$$\begin{aligned} z_T^{i,t+1} - z_T^{i,t} = (1 - z_T^{i,t})[\lambda_i z_T^{i,t} s_T^i + (1 - \lambda_i) z_T^{j,t} s_T^{ij}] - \\ - z_T^{i,t}[\lambda_i(1 - z_T^{i,t}) s_B^i + (1 - \lambda_i)(1 - z_T^{j,t}) s_B^{ij}] \end{aligned} \tag{6}$$

Here the first term represents the total probability of a B player of type i observing a T player of either type and switching to T, while the second term similarly represents a T player of type i observing a B player of either type and switching to B.

It is simple to verify that for the edge case of all players being of one type, i.e. $\lambda_i = 1$, the population dynamics coincide with the homogeneous case (2):

$$z_T^{i,t+1} - z_T^{i,t} = z_T^{i,t}(1 - z_T^{i,t})[s_T^i - s_B^i]$$

We will see in the next section that improving rules cannot exist in the heterogeneous setting. We will then focus our analysis on homogeneously dominant rules. See appendix A.2 for the population dynamics of the four fundamental homogeneously dominant rules in the heterogeneous setting.

4.1 Derivatives of the population dynamics

We form the partial derivatives of the population dynamics (5).

$$\partial(z_T^{i,t+1} - z_T^{i,t})/\partial z_T^{i,t} = \lambda_i(1 - 2z_T^{i,t})[s_T^i - s_B^i] - (1 - \lambda_i)[z_T^{j,t} s_T^{ij} + (1 - z_T^{j,t}) s_B^{ij}]$$

$$\partial(z_T^{i,t+1} - z_T^{i,t})/\partial z_T^{j,t} = (1 - \lambda_i)[(1 - z_T^{i,t}) s_T^{ij} + z_T^{i,t} s_B^{ij}]$$

We observe that depending on parameter values, the partial derivative with respect to the own type share can be either positive or negative. By contrast, the partial derivative with respect to the player proportion of the other type will always be positive. Further, both partial derivatives are continuous in $z_T^{i,t}$ and $z_T^{j,t}$. We will use these results later.

5 Non-existence of strictly improving rules

We now prove a first result:

Proposition 1. *There does not exist an improving rule with $\sigma > 0$ in the heterogeneous setting.*

Proof. By contradiction: Suppose a strictly improving rule exists in the heterogeneous setting. Then it can never give lower expected payoff than Never Switch (following Schlag (1998); see section 3.3 here). Suppose the population is initially in state $(z_T^{i,t}, z_T^{j,t}) = (0, 1)$. Then the population dynamics (5) imply for the next round:

$$\begin{aligned} z_T^{i,t+1} - z_T^{i,t} &= \lambda_i \cdot 0 \cdot (1-0)[s_T^i - s_B^i] + (1-\lambda_i)[(1-0) \cdot 1 \cdot s_T^{ij} - 0 \cdot (1-1) s_B^{ij}] = \\ &= (1-\lambda_i) s_T^{ij} \end{aligned}$$

And similarly:

$$\begin{aligned} z_T^{j,t+1} - z_T^{j,t} &= (1-\lambda_i) \cdot 1 \cdot (1-1)[s_T^j - s_B^j] + \lambda_i[(1-1) \cdot 0 \cdot s_T^{ji} - 1 \cdot (1-0) s_B^{ji}] = \\ &= -\lambda_i s_B^{ji} \end{aligned}$$

The expected switching probabilities of Never Switch are $s_T^{kl} = s_B^{kl} = 0$ for all $k, l \in I$, implying that every point is stationary under this rule. Hence, payoff in the next round will be equal to payoff in this round under this rule.

All expected switching probabilities s_C^{kl} with $C \in A$ and $k, l \in I$ of any strictly improving rule are positive for all parameter combinations $(\chi_i, \mu_i, \chi_j, \mu_j) \in (0,1)^4$. We can verify this by considering the representation of s_C^{kl} in (4). A switching rate of $\sigma > 0$ implies that $b > c \geq 0$, which in turn implies that $s_C^{kl} > 0$ for all of the above parameter configurations. Therefore $(z_T^{i,t}, z_T^{j,t}) = (0,1)$ is not stationary under any strictly improving rule for all such parameter configurations. Hence, the population must for both types in the next round move away from the corner $(z_T^{i,t}, z_T^{j,t}) = (0,1)$, following the above equations. However, if we suppose that $\chi_i < \mu_i$ and $\chi_j > \mu_i$, then expected payoff must decrease for both types as for both expected payoff is maximized in the corner $(0,1)$. Thus we have found a decision problem in which every behavioral rule with $\sigma > 0$ yields less payoff than Never Switch. Hence we have arrived at a contradiction and have proven the proposition. $\square$

Proposition 1 implies that we cannot select an optimal behavioral rule in a way similar to Schlag (1998). Instead, our approach in this document is to investigate how the optimal rules from the homogeneous setting perform in the heterogeneous setting.

We continue by studying the dynamics of our model under homogeneously improving and/or dominant rules and then evaluate them in terms of regret in the generalized heterogeneous setting. We will pay particular attention to the fundamental dominant rules of the homogeneous setting.

6 Stationarity

We are interested where the population converges to for different parameter combinations and behavioral rules. Proposition 6.3 in Weibull (1997) tells us that any limit state must be stationary.[7] Thus we first identify in this section all such stationary points or rest points for a given parameter combination and behavioral rule and then assess in section 7 the stability of these points and where the population converges to.

We find that the symmetric corners are rest points for any choice of parameters and any imitation rule. For homogeneously dominant rules under aligned preferences no other rest point exists. Under opposed preferences an

[7] We show in section 7 that the assumptions of this proposition hold in our model.

additional rest point can exist depending on the rule employed and whether certain conditions are met. We derive the general coordinates of such a rest point under PRR.

We now make the following definitions:

Definition 1. *We speak of 'aligned preferences' whenever the types prefer the same alternative, i.e. when for $C, D \in A$ with $C \neq D$ it holds that $C \succ_k D$ for all $k \in I$.*

Definition 2. *We speak of 'opposed preferences' whenever the types prefer different alternatives, i.e. when for $C, D \in A$ with $C \neq D$ it holds that $C \succ_i D$ and $D \succ_j C$ for $i, j \in I$ with $i \neq j$.*

Note that the homogeneous case is a special case of aligned preferences in which both types have the same valuation for either alternative, i.e $\chi_i = \chi_j$ and $\mu_i = \mu_j$. Further note that neither under aligned nor under opposed preferences are types allowed to be indifferent between the alternatives, i.e. $T \sim_i B \Leftrightarrow \chi_i = \mu_i$. We do not consider the case of indifference within the scope of this document.

A rest point in our setting is a root of the population dynamics (5), i.e. it must hold that $z_T^{i,t+1} = z_T^{i,t} =: z_T^i$ for both types. Recall the population dynamics:

$$z_T^{i,t+1} - z_T^{i,t} = \lambda_i z_T^{i,t}(1 - z_T^{i,t})[s_T^i - s_B^i] + (1 - \lambda_i)[(1 - z_T^{i,t}) z_T^{j,t} s_T^{ij} - z_T^{i,t}(1 - z_T^{j,t}) s_B^{ij}]$$

Then a rest point must solve the following system of equations:

$$\begin{aligned} 0 &= \lambda_i z_T^i(1 - z_T^i)[s_T^i - s_B^i] + (1 - \lambda_i)[(1 - z_T^i) z_T^j s_T^{ij} - z_T^i(1 - z_T^j) s_B^{ij}] \\ 0 &= (1 - \lambda_i) z_T^j(1 - z_T^j)[s_T^j - s_B^j] + \lambda_i[(1 - z_T^j) z_T^i s_T^{ji} - z_T^j(1 - z_T^i) s_B^{ji}] \end{aligned} \tag{7}$$

We immediately make a first observation:

Proposition 2. *The points $(z_T^i, z_T^j) = (0, 0)$ and $(z_T^i, z_T^j) = (1, 1)$ are rest points for any choice of parameters $(\chi_i, \mu_i, \chi_j, \mu_j)$ and for any imitating rule*[8].

Proof. The statement can be verified by inserting the coordinates of either point into the system of equations (7). □

[8] Recall that the population dynamics (5) have been derived for imitating rules only.

However, whether these points are also limit points and stable does depend on the parameter configuration, as we will see in section 7.

We further observe that the system of equations (7) is hard to solve. To prove that under aligned preferences there cannot exist another rest point, we therefore employ a different approach than solving the system directly.

6.1 Aligned preferences

We first prove two lemmata.

Lemma 1. *If the types play a homogeneously dominant rule and have aligned preferences with $C \succ_k D$, then for the expected switching probabilities it holds that $s_C^{ik} > s_D^{ik}$ for all $i, k \in I$.*

Proof. We assume w.l.o.g. that $T \succ_l B$ for both $l \in \{i, j\}$ and start from the general expected switching probabilities of type i being matched with type k as presented in (4):

$$\begin{aligned} s_T^{ik} - s_B^{ik} = & \mu_i \chi_k a + \mu_i (1-\chi_k) c + (1-\mu_i)\chi_k b + (1-\mu_i)(1-\chi_k) d - \\ & - [\mu_k \chi_i a + \mu_k (1-\chi_i) b + (1-\mu_k)\chi_i c + (1-\mu_k)(1-\chi_i) d] = \\ = & \quad \dots \quad = \\ = & [\chi_i - \mu_i][\mu_k (b-a) + (1-\mu_k)(d-c)] + \\ & + [\chi_k - \mu_k][\mu_i (a-c) + (1-\mu_i)(b-d)] \end{aligned}$$

Note that $T \succ_l B \Leftrightarrow \chi_l - \mu_l > 0$. With $b = 1$ and $c = 0$ for all homogeneously dominant rules this implies that none of the terms in the last expression on the RHS above can be negative. Further, the RHS has to be positive for two reasons. First, for any of the four fundamental dominant rules with $(a, d) \in \{0, 1\}^2$ at least one of the two terms will be non-zero. Second, we have argued in section 3.3 that all homogeneously dominant rules are convex combinations of the four fundamental dominant rules. Thus it follows that $s_T^{ik} - s_B^{ik} > 0$, concluding the proof. $\square$

This means in particular that $s_C^{ij} - s_D^{ij} > 0$, i.e. that the expected net switching probability when being matched with a player of the other type will be positive. Note that otherwise in general only $s_C^i - s_D^i > 0$ has to hold for homogeneously improving/dominant rules.

Lemma 2. *If the types play a homogeneously dominant rule and have aligned preferences with $C \succ_k D$, then for any point $(z_C^{i,t}, z_C^{j,t}) \in [0,1]^2/\{(0,0),(1,1)\}$ it holds that $z_C^{i,t+1} > \min(z_C^{i,t}, z_C^{j,t})$ as well as $z_C^{j,t+1} > \min(z_C^{i,t}, z_C^{j,t})$. In words: the minimum of the preferred option increases strictly.*

Proof. We assume w.l.o.g. that $T \succ_k B$ for all $k \in I$. For ease of readability we further introduce the shorthand notation $(z_T^{i,t}, z_T^{j,t}) = (x, y)$ and assume w.l.o.g. that $z_T^{i,t}$ is the minimum, i.e. $z_T^{i,t} \leq z_T^{j,t} \Leftrightarrow x \leq y$. Then we immediately see that $z_T^{i,t+1} > z_T^{i,t}$ by considering the population dynamics from (5):

$$\begin{aligned} z_T^{i,t+1} - z_T^{i,t} &= \lambda_i x(1-x)[s_T^i - s_B^i] + (1-\lambda_i)[(1-x)y s_T^{ij} - x(1-y)s_B^{ij}] \geq \\ &\geq \lambda_i x(1-x)[s_T^i - s_B^i] + (1-\lambda_i)[(1-x)x s_T^{ij} - x(1-x)s_B^{ij}] = \\ &= \lambda_i x(1-x)[s_T^i - s_B^i] + (1-\lambda_i)x(1-x)[s_T^{ij} - s_B^{ij}] > \\ &> 0 \end{aligned}$$

Here we used that $x \leq y$ implies $(1-x)y \geq (1-x)x$ and $x(1-y) \leq x(1-x)$. Further $T \succ_k B$ implies $s_T^k > s_B^k$ and $T \succ_i B \wedge T \succ_j B$ implies $s_T^{ij} > s_B^{ij}$ following Lemma 1.

Thus it must hold that $z_T^{i,t+1} > z_T^{i,t}$.

It then remains to be shown that $z_T^{j,t+1} > z_T^{i,t}$. We rearrange the population dynamics for type j as follows:

$$\begin{aligned} z_T^{j,t+1} - z_T^{j,t} &= (1-\lambda_i)y(1-y)[s_T^j - s_B^j] + \lambda_i[(1-y)x s_T^{ji} - y(1-x)s_B^{ji}] = \\ &= (1-\lambda_i)y(1-y)[s_T^j - s_B^j] + \lambda_i y(1-y)[s_T^{ji} - s_B^{ji}] - \\ &\quad - \lambda_i(y-x)[(1-y)s_T^{ji} + y s_B^{ji}] \end{aligned}$$

We now show that $z_T^{j,t+1} > z_T^{i,t}$:

$$\begin{aligned} z_T^{j,t+1} - z_T^{i,t} &= z_T^{j,t+1} - z_T^{j,t} + z_T^{j,t} - z_T^{i,t} = \\ &= z_T^{j,t+1} - z_T^{j,t} + y - x = \\ &= (1-\lambda_i)y(1-y)[s_T^j - s_B^j] + \lambda_i y(1-y)[s_T^{ji} - s_B^{ji}] - \\ &\quad - \lambda_i(y-x)[(1-y)s_T^{ji} + y s_B^{ji}] + y - x = \\ &= (1-\lambda_i)y(1-y)[s_T^j - s_B^j] + \lambda_i y(1-y)[s_T^{ji} - s_B^{ji}] + \\ &\quad + (y-x)\left[1 - \lambda_i[(1-y)s_T^{ji} + y s_B^{ji}]\right] > \\ &> 0 \end{aligned}$$

The inequality follows as the first two terms on the RHS are positive and the third non-negative. The reason for the first two terms is that $s_T^j > s_B^j$ and $s_T^{ji} > s_B^{ji}$ as we just argued above. For the third term it follows from $y \geq x$ and as it must hold that $\lambda_i[(1-y)s_T^{ji} + ys_B^{ji}] < 1$.

We have now proven that the minimum of the probability of playing the better alternative among the types will increase in every point (except the symmetric corner points), i.e. that $z_T^{i,t+1} > \min(z_T^{i,t}, z_T^{j,t})$ and $z_T^{j,t+1} > \min(z_T^{i,t}, z_T^{j,t})$. $\square$

We can now prove the following proposition:

Proposition 3. *If the types play a homogeneously dominant rule and have aligned preferences, then the only rest points are the symmetric corners* $(z_T^i, z_T^j) = (0,0)$ *and* $(z_T^i, z_T^j) = (1,1)$.

Proof. The statement follows immediately from Lemma 2. As the minimum of the preferred option increases strictly in all points except for the symmetric corners, no point other than those corners can be a rest point. $\square$

6.2 Opposed preferences

To find conditions for the existence of other rest points under opposed preferences we have to turn back to the system of quadratic equations (7). We limit our analysis in this subsection to the rule PRR, as it allows us to simplify the system. We have already seen in section 3.4 that for this and only this rule $s_C^{ij} = s_C^i$ for all $C \in A$. The reason for this equality is once again that PRR does not take the payoff of the observed other into consideration when determining the action for the next round.
Thus we solve the following system of equations:

$$0 = \lambda_i z_T^i(1-z_T^i)[s_T^i - s_B^i] + (1-\lambda_i)[(1-z_T^i)z_T^j s_T^i - z_T^i(1-z_T^j)s_B^i]$$
$$0 = (1-\lambda_i)z_T^j(1-z_T^j)[s_T^j - s_B^j] + \lambda_i[(1-z_T^j)z_T^i s_T^j - z_T^j(1-z_T^i)s_B^j]$$

The solution given by equation (8) can only exist under the following conditions. For once, it must always hold that $z_T^i, z_T^j \in [0,1]$. Further, the denominator can never be zero, which implies that $s_T^i \neq s_B^i$ for both types as well as $s_T^i s_B^j \neq s_B^i s_T^j$. Our definition of opposed preferences already guarantees that the first condition must hold. Note that the second condition

is equivalent to $(s_T^i - s_B^i)s_B^j \neq s_B^i(s_T^j - s_B^j)$. Hence this condition will also always hold as the sides must have a different sign per definition of opposed preferences.
Given that the condition $z_T^i, z_T^j \in [0,1]$ holds, the solution to the above system of equations is given by:

$$z_T^i = \frac{\lambda_i[s_T^i - s_B^i]s_B^j + (1-\lambda_i)[s_T^j - s_B^j]s_B^i}{\lambda_i[s_T^i - s_B^i][s_T^i s_B^j - s_B^i s_T^j]} s_T^i \tag{8}$$

This identity must hold equivalently for type j, as the problem is symmetric for the types. Notice that the enumerator has the same value for both types. We will derive simpler conditions that guarantee that $z_T^i, z_T^j \in [0,1]$ in Proposition 4. First, we now derive the representation of an internal rest point under PRR by inserting for the expected switching probabilities according to Table 1:

$$z_T^i = \frac{\lambda_i[\chi_i - \mu_i](1-\chi_j) + (1-\lambda_i)[\chi_j - \mu_j](1-\chi_i)}{\lambda_i[\chi_i - \mu_i][(1-\mu_i)(1-\chi_j) - (1-\chi_i)(1-\mu_j)]}(1-\mu_i) \tag{9}$$

For the sake of clarity we will use a simplified representation based on (8) in the upcoming analysis. As PRR is a homogeneously dominant rule, it holds that $s_T^i - s_B^i = \sigma(\chi_i - \mu_i)$ with $\sigma = 1$. Using this identity we simplify (8) to:

$$z_T^i = \frac{\lambda_i[\chi_i - \mu_i]s_B^j + (1-\lambda_i)[\chi_j - \mu_j]s_B^i}{\lambda_i[\chi_i - \mu_i][s_T^i s_B^j - s_B^i s_T^j]} s_T^i \tag{10}$$

An equivalent representation is:

$$z_T^i = \frac{\lambda_i[\chi_i - \mu_i]s_T^i s_B^j + (1-\lambda_i)[\chi_j - \mu_j]s_B^i s_T^i}{\lambda_i[\chi_i - \mu_i]s_T^i s_B^j - \lambda_i[\chi_i - \mu_i]s_B^i s_T^j} \tag{11}$$

We now prove the following result, which gives us simplified conditions that guarantee that $z_T^i, z_T^j \in [0,1]$.

Proposition 4. *If the types play the rule PRR and have opposed preferences, then an additional rest point exists at the coordinates given by* (8)*, if the following conditions hold:*

$$\lambda_i[\chi_i - \mu_i]s_B^j + (1-\lambda_i)[\chi_j - \mu_j]s_B^i > 0 \tag{12}$$

$$\lambda_i[\chi_i - \mu_i]s_T^j + (1-\lambda_i)[\chi_j - \mu_j]s_T^i < 0 \tag{13}$$

Proof. We assume w.l.o.g. that $\chi_i - \mu_i > 0$, $\chi_j - \mu_j < 0$ and $s_T^i s_B^j - s_B^i s_T^j > 0$. Then from (10) it follows that:

$$z_T^i = \frac{\lambda_i[\chi_i - \mu_i]s_B^j + (1-\lambda_i)[\chi_j - \mu_j]s_B^i}{\lambda_i[\chi_i - \mu_i][s_T^i s_B^j - s_B^i s_T^j]} s_T^i > 0$$
$$z_T^j = \frac{\lambda_i[\chi_i - \mu_i]s_B^j + (1-\lambda_i)[\chi_j - \mu_j]s_B^i}{(1-\lambda_i)[\chi_j - \mu_j][s_T^j s_B^i - s_B^j s_T^i]} s_T^j > 0$$

Both inequalities simplify to (12) as above.

Further it follows from (11) that[9]:

$$\begin{aligned}
z_T^i &< 1 \\
\frac{\lambda_i[\chi_i - \mu_i]s_T^i s_B^j + (1-\lambda_i)[\chi_j - \mu_j]s_B^i s_T^i}{\lambda_i[\chi_i - \mu_i]s_T^i s_B^j - \lambda_i[\chi_i - \mu_i]s_B^i s_T^j} &< 1 \\
\lambda_i[\chi_i - \mu_i]s_T^i s_B^j + (1-\lambda_i)[\chi_j - \mu_j]s_B^i s_T^i &< \lambda_i[\chi_i - \mu_i]s_T^i s_B^j - \lambda_i[\chi_i - \mu_i]s_B^i s_T^j \\
(1-\lambda_i)[\chi_j - \mu_j]s_B^i s_T^i &< -\lambda_i[\chi_i - \mu_i]s_B^i s_T^j \\
\lambda_i[\chi_i - \mu_i]s_T^j + (1-\lambda_i)[\chi_j - \mu_j]s_T^i &< 0
\end{aligned}$$

Again the same inequality (13) follows from $z_T^j < 1$.

Thus we have proven that a third rest point at the coordinates given by (8) only exists under opposed preferences and if conditions (12) and (13) are met. □

With Propositions 2-4 we have now derived the rest points under aligned preferences and under the PRR with opposed preferences. We will use them to study the convergence behavior of the population in section 7. Before we continue we will consider stationarity and population behavior on the edges of the $z_T^t \in [0,1]^2$ area.

[9]Here every line is equivalent to the one above it. Note that in the step from the second to the third line the inequality sign does not change direction, as the denominator is positive given our assumptions w.l.o.g. at the start of the proof.

6.3 Edges

Recall the population dynamics (5):

$$z_T^{i,t+1} - z_T^{i,t} = \lambda_i z_T^{i,t}(1-z_T^{i,t})[s_T^i - s_B^i] + (1-\lambda_i)[(1-z_T^{i,t})z_T^{j,t}s_T^{ij} - z_T^{i,t}(1-z_T^{j,t})s_B^{ij}]$$

On the edge $(z_T^{i,t}, z_T^{j,t}) = (0, y)$ with $y \in (0,1)$:

$$\begin{aligned} z_T^{i,t+1} - z_T^{i,t} &= \lambda_i \cdot 0 \cdot (1-0)[s_T^i - s_B^i] + (1-\lambda_i)[(1-0)ys_T^{ij} - 0 \cdot (1-y)s_B^{ij}] = \\ &= (1-\lambda_i)ys_T^{ij} \end{aligned}$$

$$\begin{aligned} z_T^{j,t+1} - z_T^{j,t} &= (1-\lambda_i)y(1-y)[s_T^j - s_B^j] + \lambda_i[(1-y) \cdot 0 \cdot s_T^{ji} - y(1-0)s_B^{ji}] = \\ &= (1-\lambda_i)y(1-y)[s_T^j - s_B^j] - \lambda_i y s_B^{ji} \end{aligned}$$

On the edge $(z_T^{i,t}, z_T^{j,t}) = (1, y)$ with $y \in (0,1)$:

$$\begin{aligned} z_T^{i,t+1} - z_T^{i,t} &= \lambda_i \cdot 1 \cdot (1-1)[s_T^i - s_B^i] + (1-\lambda_i)[(1-1)ys_T^{ij} - 1 \cdot (1-y)s_B^{ij}] = \\ &= -(1-\lambda_i)(1-y)s_B^{ij} \end{aligned}$$

$$\begin{aligned} z_T^{j,t+1} - z_T^{j,t} &= (1-\lambda_i)y(1-y)[s_T^j - s_B^j] + \lambda_i[(1-y) \cdot 1 \cdot s_T^{ji} - y(1-1)s_B^{ji}] = \\ &= (1-\lambda_i)y(1-y)[s_T^j - s_B^j] + \lambda_i(1-y)s_T^{ji} \end{aligned}$$

By symmetry equivalent results hold for the remaining two edges.

We observe that the edges are not stationary for any combination of decision rules and parameter configurations that results in a positive corresponding expected switching probability $s_C^{ik} > 0$ for all $C \in A$ and $k \in I$. For homogeneously improving rules that is the case for all parameter configurations $(\chi_i, \mu_i, \chi_j, \mu_j) \in (0,1)^4$. However, for extreme parameter configurations a rest point according to (8) can exist on the edges. For homogeneously improving rules such a configuration would be extreme if the expected payoff of at least one alternative and at least one type must be either 0 or 1.

Further we see that for any decision rule and parameter configuration the population stays on the edge or moves back towards the interior of the area $(z_T^{i,t}, z_T^{j,t}) \in [0,1]^2$, but never leaves it. We will use this result in the proof of Proposition 5.

7 Convergence and stability

Having derived the stationary points for aligned preferences and for opposed preferences under PRR in the last section, we continue our analysis of the convergence behavior of the population in this section.

In numerical simulations we observe that for every 'well behaved'[10] parameter combination, homogeneously improving rule and starting point there will exist one globally internally asymptotically stable point, i.e. the population will converge towards that point from any starting point within $[0,1]^2/\{(0,0),(1,1)\}$. However, we have only found an analytical proof for aligned preferences. We will study the behavior under opposed preferences using numerical simulation.

7.1 Aligned preferences

As is stated in Weibull (1997), a common method of proving stability of or convergence behavior towards a stationary point is the direct Lyapunov method. However, in our case it proves difficult to find a suitable Lyapunov function. The individual type populations as well as the total population average will in general not move montonously towards the preferred symmetric corner and are thus not suitable as a Lyapunov function.[11]

In the following proof we do not employ the direct Lyapunov method, but instead use the result of Lemma 2 that the minimum of the share of the preferred alternative among both types must increase strictly. In a second step we then prove that the population thus must converge towards the preferred symmetric corner.

We note here that this minimum is a suitable Lyapunov function. However, the limit state in question is located at the corner of the $[0,1]^2$ area of *relevant* states, i.e. of those states that have a meaningful interpretation in our model. This implies that we would have to consider the dynamic properties of our system given by the population dynamics (5) outside this area of relevant states, as Theorem 6.2 in Weibull (1997) requires the Lyapunov function to have its characteristic properties in a neighborhood of the limit state. As

[10] We exclude 'extreme' valuation parameters (e.g. $\chi_i \in \{0,1\}$) and the symmetric corners as starting points.

[11] As a counterexample consider $(\chi, \mu_i, \chi_j, \mu_j) = (0.2, 0.1, 0.9, 0.8)$ with $\lambda_i = 0.5$ in state $z_T^t = (0,1)$ under PRR.

we do not want to consider any $z_T^{i,t} \notin [0,1]$ - as would be required by such a proof -, we instead conduct the proof as described above and detailed in the following.

Proposition 5. *If the types play a homogeneously dominant rule and have aligned preferences, then they will both converge towards playing their preferred alternative with certainty.*

Proof. We assume w.l.o.g. that $T \succ_k B$ for $k \in \{i,j\}$. Thus we have to prove that the population will converge towards the corner $(z_T^{i,t}, z_T^{j,t}) = (1,1)$.

Given the above assumptions, we know from Lemma 2 that the minimum of z_T^t must increase strictly. This implies that the state variable z_T^t must converge towards one or more limit points. Proposition 6.3 in Weibull (1997) tells us that limit points must be rest points, provided certain additional assumptions apply. The only rest points under aligned preferences are the symmetric corners, as we have proven in Proposition 3. Given that the minimum of z_T^t increases strictly, z_T^t must converge towards the corner $(1,1)$ if the assumptions for Proposition 6.3. apply.

We demonstrate in appendix A.3 that the assumptions for Proposition 6.3 do apply and thus the population will converge towards the corner $(z_T^{i,t}, z_T^{j,t}) = (1,1)$. $\square$

From Proposition 5 it immediately follows that under the same condition $T \succ_k B$ the T corner $z_T^t = (1,1)$ is globally internally asymptotically stable following Definition 6.5 in Weibull (1997).
The T corner is Lyapunov stable as for every neighborhood B of it we can construct a neighborhood B^0 of the T corner with $B^0 = \{(z_T^i, z_T^j) \in C \mid 1 - z_T^i \leq \varepsilon \wedge 1 - z_T^j \leq \varepsilon\}$ where $\varepsilon > 0$ is small enough such that $B^0 \subset B$. Then every population state $\xi(t, x^0)$ of a point $x^0 \in B^0$ must necessarily lie within $B^0 \cap C$ for all $t \geq 0$ (as we have shown that the minimum must increase) and thus also within B.
The T corner is further globally internally asymptotically stable as we have already shown that all points in $C/\{(0,0)\}$ converge to it. The basin of attraction of the T corner is $C/\{(0,0)\}$. We exclude the other symmetric corner as we have shown that it is stationary.

7.2 Opposed preferences

As we have stated at the beginning of this section, we only describe the population behavior under opposed preferences from observations in numerical simulations.

We observe in the simulations that if an interal rest point exists according to Proposition 4, then it will be globally internally asymptotically stable, i.e the population will converge there from any starting point except the symmetric corners.

However, if one of the conditions in Proposition 4 does not hold and an internal rest point hence does not exist, then the population will converge towards one of the symmetric corners. We can determine to which of the two symmetric corners they will converge based on which condition is violated:[12] If condition (12) is violated, then the population converges towards the B corner. Conversely, if condition (13) is violated, then the population converges towards the T corner.

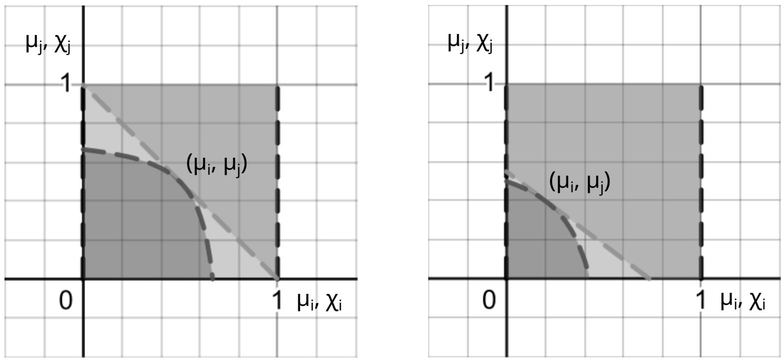
We note here that even if one of the conditions in Proposition 4 is violated and the population thus converges towards a symmetric corner, we cannot use the same argument as in Proposition 5. The reason is that the minimum does not necessarily increase for opposed preferences.[13]

7.3 Illustration

We illustrate our findings by visualizing the different areas of convergence under PRR. Fig. 2 shows them for different parameter values. λ_i is fixed and not depicted. μ_i and μ_j are fixed and depicted. The movement along the axes represents change in the parameters χ_i and χ_j respectively. Note that type i prefers T on the right half of the $[0,1]^2$ area and B on the left half. Type j prefers T on the top half and B on the bottom half. Preferences are thus aligned in the top right and bottom left quadrant of the $[0,1]^2$ area and are opposed in the top left and bottom right quadrant. The dark grey and medium grey areas correspond to parameter conditions for which either

[12] It can be shown that at least one condition must hold at any time.

[13] As a counterexample consider $(\chi, \mu_i, \chi_j, \mu_j) = (0.4, 0.5, 0.9, 0.5)$ with $\lambda_i = 0.5$ in state $z_T^t = (0.5, 0.5)$ under PRR.



(a) PRR, $(\mu_i, \mu_j) = (0.5, 0.5)$ (b) PRR, $(\mu_i, \mu_j) = (0.2, 0.4)$

Figure 2: Different areas of convergence under PRR for $\lambda_i = 0.5$. The (x, y) coordinates correspond to (χ_i, χ_j) while (μ_i, μ_j) is fixed - on the left to $(0.5, 0.5)$ and on the right to $(0.2, 0.4)$. The bottom left dark grey area corresponds to parameter combinations that result in a convergence towards the B option respectively and the top right medium grey area similarly to the T option. In the light grey areas close to the $(0, 1)$ and $(1, 0)$ corners the population converges towards an internal point with coordinates according to (8).

of the conditions for the existence of an internal rest point of Proposition 4 is violated. The population converges towards the B corner in the dark grey and towards the T corner in the medium grey area. We observe that in general these are areas in which one type values their preferred alternative relatively stronger than the other values theirs. In the light grey area the population converges towards an internal rest point as given by (8).

To conclude this section we consider how the value of λ_i affects convergence behavior. The proof of Proposition 5 for aligned preferences is valid for all values of λ_i. However, for opposed preferences under PRR the two conditions for the existence of an internal rest point depend on λ_i. We observe that if λ_i is skewed far enough towards one type, then one of the conditions will not hold and the population will converge towards the corner preferred by the type that now has a higher share of the total population. This is what we would intuitively expect to be the case. The value of λ_i for which this will be the case depends on the parameter configuration.

8 Optimal rules

Having understood the dynamics of our model, we can now evaluate the performance of different homogeneously dominant rules in this setting to find the optimal rule(s).
In this section we assess their performance in terms of minimax regret. The optimal rule under minimax regret, as first introduced in Savage (1951), will be the rule that has the least regret in their worst case. We investigate a short-run version of this measure by minimizing the maximum regret to the next round. We restrict our analysis in this section to homogeneously dominant rules.

Regret for type i at any point $z_T^t \in [0,1]^2$ is given by:

$$r_i = \begin{cases} z_B^{i,t}(\chi_i - \mu_i) & \text{if } T \succ_i B \\ z_T^{i,t}(\mu_i - \chi_i) & \text{if } B \succ_i T \\ 0 & \text{else} \end{cases} \tag{14}$$

Note that in this section we only need to examine either the case $T \succ_i B$ or $B \succ_i T$, as all behavioral rules we are considering in this document prescribe the same behavior if alternatives were switched and so the worst cases in terms of regret will be symmetric for both.[14]

8.1 Regret to next round

We now assess regret from one round to the next. We derive maximum regret for an arbitrary homogeneously dominant rule and identify the rules that minimize this maximum regret.[15]

[14] See also the discussion in section 3.3 after the introduction of a short-hand notation for the parameters of imitation rules for extreme payoffs.

[15] We do not search among all homogeneously improving rules, as we can immediately infer that Never Switch will be the homogeneously improving rule that minimizes this maximum regret, as its regret to next round is always zero. If we exclude such degenerate rules then there does not exist a homogeneously improving rule that minimizes maximum regret, because the regret to next round can be made arbitrarily small by lowering the expected switching probabilities by decreasing the parameters (a, b, c, d).

Following the general population dynamics the change in population from round t to $t+1$ is equal to $z_T^{i,t+1} - z_T^{i,t}$. Then the regret to next round is:

$$r_i^{t,t+1} = \max\left[0, -(z_T^{i,t+1} - z_T^{i,t})(\chi_i - \mu_i)\right]$$

Note that $r_i^{t,t+1}$ is a function of the decision rule parameters (a,b,c,d), of the valuation parameters $(\mu_i, \mu_j, \chi_i, \chi_j)$, of the type share λ_i and of the shares of T players $z_T^{i,t}$, $z_T^{j,t}$ and $z_T^{i,t+1}$ respectively.
The above expression can be understood such that the population regrets the change if it moves them away from their preferred option and otherwise does not regret it. Note that this expression holds for both the case where T is the better option as well as the case where B is the better option.[16]

We look for the parameter configuration that maximizes this regret. The parameters are $z_T^{i,t}, z_T^{j,t}, \chi_i, \mu_i, \chi_j, \mu_j$ and λ_i. Notice that $z_T^{i,t+1}$ is determined entirely by these parameters. We note at this point that $z_T^{i,t}$ is an unknown parameter for every member of the population, as this is the actual proportion of players from population i that play T, which is unknown to every individual member (who plays in every round their current strategy depending on their realized last round). We assume that maximum regret is positive and will therefore omit the comparison with zero in the following. We introduce the shorthand notation:

$$\max r_i^{t,t+1} := \max_{\substack{z_T^{i,t}, z_T^{j,t}, \lambda_i, \\ \chi_i, \mu_i, \chi_j, \mu_j}} r_i^{t,t+1}$$

Once we have attained maximum regret for all possible homogeneously dominant rules (i.e. for all combinations $(a,b,c,d) \in [0,1]^4$) we can identify the rules that minimize it.
Thus we are solving the following min-maximization problem:

$$\min_{a,b,c,d} \max_{\substack{z_T^{i,t}, z_T^{j,t}, \lambda_i, \\ \chi_i, \mu_i, \chi_j, \mu_j}} r_i^{t,t+1}$$

See appendix A.4 for the derivation of the solution. We there find that the homogeneously dominant rules that minimize maximum regret are all convex combinations of ρ - PIR and PRR. Maximum regret of these rules is $\max r_i^{t,t+1} = 0.25$.

[16] $-(z_T^{i,t+1} - z_T^{i,t})(\chi_i - \mu_i) = (1 - z_T^{i,t+1} - (1 - z_T^{i,t}))(\chi_i - \mu_i) = (z_B^{i,t+1} - z_B^{i,t})(\chi_i - \mu_i)$

We intuitively make sense of the parameter combination for which this maximum regret is attained. As we assume in the derivation that $T \succ_i B$, the worst case is attained at $z_T^t = (1, 0)$, i.e. if all players of the own type play T and all players of the other type play B. Further $\lambda_i \to 0$ - this can be understood as there being no other player of the own type, maximizing the probability of switching to B by only meeting those players. For the valuations we find $\chi_i = 0.5, \mu_i = 0, \chi_j \in [0, 1]$. The regret maximizing value of μ_j depends on the rule used - for PRR all values $\mu_j \in [0, 1]$ are possible as that rule does not use the information from the other's payoff. For all other rules it is $\mu_j = 1$. Unsurprisingly, this value maximizes the probability of switching to B. The value of $\chi_i = 0.5$ can be understood in the following way - a lower value of χ_i makes it more probable that the player switches away to their less preferred alternative (here B), while a higher value increases the payoff lost from such a switch. Both contribute to a higher regret and so the maximizing value lies in the middle.

9 Conclusion

In this document we have studied the generalization of the social learning model presented in Schlag (1998) from a homogeneous setting, where all individuals in the population have the same preferences, to a heterogeneous setting in which there exist two types with different preferences. In this section we recapitulate our findings and discuss possible extensions.

We have found that in this generalized setting there no more exist any *improving rules* - simple behavioral rules that guarantee a convergence towards the preferred alternative. We then have focused our analysis on the 'best' improving rules from the homogeneous setting - the dominant rules. We have further limited our analysis to dominant rules for extreme payoffs.
For those rules we have found that if preferences are *aligned*, i.e. both types prefer the same alternative, then both types will in the limit play that alternative with certainty under any such rule. If preferences are instead *opposed*, i.e. the types prefer different alternatives, then we have found that under the dominant rule PRR there exist conditions under which they will instead in the limit play a mixed action and not converge towards playing either alternative with certainty. As we have in the scope of this document only derived the behavior under PRR, a natural extension here is to study the

convergence behavior under the other dominant rules.
Finally we have determined the rules that perform best given a certain metric. We have chosen this metric to be minimax regret to the next round. We have found that both the extreme payoff equivalent to PIR as well as PRR perform best in regard to this metric.[17] A possible extension here would be to chose a different metric, for example minimax regret at the convergence point, which would represent a more long-run metric.

We conclude by discussing two further possible extensions. The first is the question how differing degrees of heterogeneity affect the performance of a rule. We have seen that regret to the next round is maximized if the types' preferences differ as strongly as possible. Therefore we would expect that a limit to the degree of heterogeneity would also improve the rules' performance in the worst case, justifying the usage of the homogeneously dominant rules PIR and PRR even in a heterogeneous setting.
The second extension pertains to the observation that numerical simulations suggest that the speed of convergence depends on the used rule. In particular, it seems that certain rules generally prescribe stronger changes than others. A topic of interest could then be whether there exists a relation between speed of convergence and the performance of a rule, for example in regard to maximum regret at the convergence point.

[17] Further any rule that is a convex combination of those two rules performs equally well.

References

Björnerstedt, Jonas and Karl H Schlag (1996). *On the evolution of imitative behavior*. Discussion Paper. University of Bonn.

Cunningham, Daniel W (2021). *Real Analysis: With Proof Strategies*. Chapman and Hall/CRC.

Kirman, Alan (2006). "Heterogeneity in economics". In: *Journal of Economic Interaction and Coordination* 1(1), pp. 89–117.

Köster, Raphael et al. (2020). "Model-free conventions in multi-agent reinforcement learning with heterogeneous preferences". In: *arXiv preprint arXiv:2010.09054*.

Savage, Leonard J (1951). "The theory of statistical decision". In: *Journal of the American Statistical association* 46(253), pp. 55–67.

Schlag, Karl H (1998). "Why imitate, and if so, how? A boundedly rational approach to multi-armed bandits". In: *Journal of economic theory* 78(1), pp. 130–156.

Schlag, Karl H (1999). "Which one should I imitate?" In: *Journal of Mathematical Economics* 31(4), pp. 493–522.

Schlag, Karl H (2004). *Imitation and Minimax Regret*. Mimeo. Accessed on October 10, 2022. URL: `http://homepage.univie.ac.at/karl.schlag/research/imitation/regretim47.pdf`.

Schlag, Karl H et al. (2013). "Eliciting probabilities, means, medians, variances and covariances without assuming risk neutrality". In: *Theoretical Economics Letters*.

Schlag, Karl H (2022). "Social Learning between Groups: Imitation and the Role of Experience". In: *Games* 13(5), p. 60.

Weibull, Jörgen W (1997). *Evolutionary game theory*. MIT press.

A Appendix

A.1 Derivation of the population dynamics

For type population i the expected share of T players in the next round can be determined by adding up the expected probabilities of type i players playing action T after they observe another player, weighted by the probability of that specific match occurring. For example, the first term on the second line on the RHS of the following equation gives the probability $F_T^{ii}(T,B)$ of a T player of type i who observes a B player of type i and who will play T again in the next round, weighted by the probability of this match occurring, which is the share z_T^i of T players of type i times the share of B players of type i in the whole population, $\lambda_i(1-z_T^i)$.

$$\begin{aligned}
z_T^{i,t+1} = {} & \lambda_i(z_T^{i,t})^2 F_T^{ii}(T,T) + (1-\lambda_i) z_T^{i,t} z_T^{j,t} F_T^{ij}(T,T) + \\
& + \lambda_i z_T^{i,t}(1-z_T^{i,t}) F_T^{ii}(T,B) + (1-\lambda_i) z_T^{i,t}(1-z_T^{j,t}) F_T^{ij}(T,B) + \\
& + \lambda_i(1-z_T^{i,t})^2 F_T^{ii}(B,B) + (1-\lambda_i)(1-z_T^{i,t})(1-z_T^{j,t}) F_T^{ij}(B,B) + \\
& + \lambda_i(1-z_T^{i,t}) z_T^{i,t} F_T^{ii}(B,T) + (1-\lambda_i)(1-z_T^{i,t}) z_T^{j,t} F_T^{ij}(B,T)
\end{aligned}$$

We only consider imitating rules, which implies:

$$F_T^{ik}(T,T) = 1 \quad \text{and} \quad F_T^{ik}(B,B) = 0$$

Further we know that:

$$F_T^{ik}(T,B) = 1 - F_B^{ik}(T,B) = 1 - s_B^{ik}$$

Thus we continue:

$$\begin{aligned}
z_T^{i,t+1} &= \lambda_i(z_T^{i,t})^2 \cdot 1 + (1-\lambda_i) z_T^{i,t} z_T^{j,t} \cdot 1 + \\
&\quad + \lambda_i z_T^{i,t}(1-z_T^{i,t})(1-s_B^i) + (1-\lambda_i) z_T^{i,t}(1-z_T^{j,t})(1-s_B^{ij}) + \\
&\quad + \lambda_i(1-z_T^{i,t})^2 \cdot 0 + (1-\lambda_i)(1-z_T^{i,t})(1-z_T^{j,t}) \cdot 0 + \\
&\quad + \lambda_i(1-z_T^{i,t}) z_T^{i,t} s_T^i + (1-\lambda_i)(1-z_T^{i,t}) z_T^{j,t} s_T^{ij} = \\
&= \lambda_i z_T^{i,t} + (1-\lambda_i) z_T^{i,t} - \\
&\quad - \lambda_i z_T^{i,t}(1-z_T^{i,t}) s_B^i - (1-\lambda_i) z_T^{i,t}(1-z_T^{j,t}) s_B^{ij} + \\
&\quad + \lambda_i(1-z_T^{i,t}) z_T^{i,t} s_T^i + (1-\lambda_i)(1-z_T^{i,t}) z_T^{j,t} s_T^{ij} = \\
&= z_T^{i,t} + \lambda_i z_T^{i,t}(1-z_T^{i,t})[s_T^i - s_B^i] + \\
&\quad + (1-\lambda_i)[(1-z_T^{i,t}) z_T^{j,t} s_T^{ij} - z_T^{i,t}(1-z_T^{j,t}) s_B^{ij}]
\end{aligned}$$

So the population dynamics are:

$$\begin{aligned}z_T^{i,t+1} = z_T^{i,t} + \lambda_i z_T^{i,t}(1 - z_T^{i,t})[s_T^i - s_B^i] + \\ + (1 - \lambda_i)[(1 - z_T^{i,t})z_T^{j,t} s_T^{ij} - z_T^{i,t}(1 - z_T^{j,t})s_B^{ij}]\end{aligned}$$

We can see that in the edge case where $\lambda_i = 1$ we obtain the known population dynamics for homogeneous preferences:

$$z_T^{i,t+1} = z_T^{i,t} + z_T^{i,t}(1 - z_T^{i,t})[s_T^i - s_B^i]$$

For the case $\lambda_i = 0$ we obtain:

$$z_T^{i,t+1} = z_T^{i,t} + (1 - z_T^{i,t})z_T^{j,t} s_T^i - z_T^{i,t}(1 - z_T^{j,t})s_B^i$$

Which does not have a meaningful interpretation, since population i does not exist for these parameters.
Similarly, if we assume $z_T^{j,t} = z_T^{i,t}$ we obtain:

$$\begin{aligned}z_T^{i,t+1} &= z_T^{i,t} + \lambda_i z_T^{i,t}(1 - z_T^{i,t})[s_T^i - s_B^i] + \\ &\quad + (1 - \lambda_i)[(1 - z_T^{i,t})z_T^{i,t} s_T^i - z_T^{i,t}(1 - z_T^{i,t})s_B^i] \\ &= z_T^{i,t} + (\lambda_i + (1 - \lambda_i))z_T^{i,t}(1 - z_T^{i,t})[s_T^i - s_B^i] \\ &= z_T^{i,t} + z_T^{i,t}(1 - z_T^{i,t})[s_T^i - s_B^i]\end{aligned}$$

Which are again the known population dynamics for homogeneous preferences. Interpretation: This is an interesting property, as it implies that if the fraction of each type population that plays T is the same in a given round t', then the population dynamics are identical to the homogeneous case. However, both types will diverge again unless they both prefer the same option by the exact same margin. Even then, the population dynamics only hold in expectation.[18]

A.2 Population dynamics of homogeneously dominant rules

We derive the population dynamics of the four fundamental homogeneously dominant rules in the heterogeneous setting. In each case we insert the expected switching probabilities as derived in section 3.4 into the population

[18] We have however assumed an infinite population, in which case they hold in practice as well.

dynamics (5) and simplify the expression.

Population dynamics of ρ - PIR:

$$z_T^{i,t+1} - z_T^{i,t} = \lambda_i z_T^{i,t}(1 - z_T^{i,t})[\chi_i - \mu_i] + \\ + (1 - \lambda_i)[(1 - z_T^{i,t}) z_T^{j,t}(1 - \mu_i)\chi_j - z_T^{i,t}(1 - z_T^{j,t})\mu_j(1 - \chi_i)]$$

Population dynamics of POR:

$$z_T^{i,t+1} - z_T^{i,t} = \lambda_i z_T^{i,t}(1 - z_T^{i,t})[\chi_i - \mu_i] + \\ + (1 - \lambda_i)[(1 - z_T^{i,t}) z_T^{j,t}\chi_j - z_T^{i,t}(1 - z_T^{j,t})\mu_j]$$

Population dynamics of PRR:

$$z_T^{i,t+1} - z_T^{i,t} = \lambda_i z_T^{i,t}(1 - z_T^{i,t})[\chi_i - \mu_i] + \\ + (1 - \lambda_i)[(1 - z_T^{i,t}) z_T^{j,t}(1 - \mu_i) - z_T^{i,t}(1 - z_T^{j,t})(1 - \chi_i)]$$

Population dynamics of ρ - RPIR:

$$z_T^{i,t+1} - z_T^{i,t} = \lambda_i z_T^{i,t}(1 - z_T^{i,t})[\chi_i - \mu_i] + \\ + (1 - \lambda_i)[(1 - z_T^{i,t}) z_T^{j,t}(1 - \mu_i(1 - \chi_j)) - z_T^{i,t}(1 - z_T^{j,t})(1 - \chi_i(1 - \mu_j))]$$

A.3 Assumptions for proof of Proposition 5

We show here that the assumptions of Proposition 6.3 apply and we thus can use its result in the proof of Proposition 5.

All numbered propositions and definitions in the remainder of this subsection can be found in Weibull (1997) unless stated otherwise.

The following assumptions for Proposition 6.3 are listed in Proposition 6.1:[19]

1. $X \subset R^k$ is open
2. $\varphi : X \to R^k$ is Lipschitz continuous
3. C is a compact subset of X such that $\xi(t, x^0) \in C$ for all $x^0 \in C$ and $t \in T(x^0)$

[19]Note that Weibull (1997) denotes $\mathbb{R}$ by R.

We now explain the notation[20] and prove that all three assumptions apply. k is the dimension of the dynamic system, in our case $k = 2$. X is the state space, in our case $X = \mathbb{R}^2$ - assumption 1 thus applies as $\mathbb{R}^n$ is open. A state vector $x \in X$ would in our case be $x = z_T^t = (z_T^{i,t}, z_T^{j,t})$.

The vector field φ gives the change in population as a function of the state vector - in our case its components are each equal to the RHS of the population dynamics (5):

$$\begin{aligned}\varphi_i(x) = \varphi_i(z_T^t) = \lambda_i z_T^{i,t}(1 - z_T^{i,t})[s_T^i - s_B^i] + \\ + (1 - \lambda_i)[(1 - z_T^{i,t})z_T^{j,t} s_T^{ij} - z_T^{i,t}(1 - z_T^{j,t})s_B^{ij}]\end{aligned}$$

We show that φ is Lipschitz continuous using Definition 6.2:
We have to show that “for every compact subset $C \subset X$ there exists some real number λ such that [the following] holds for all $x, y \in C$: $\|\varphi(x) - \varphi(y)\| \leq \lambda\|x - y\|$”.
We make use of the remark that is following this definition in Weibull (1997), which states that it can be shown that continuous first partial derivatives of φ imply Lipschitz continuity. We denote the partial derivatives of φ by $D_i\varphi$ and $D_j\varphi$. We have already argued in section 4.1 that they are continuous. Then by the extreme value theorem[21] the partial derivatives have to attain their maximum on a compact subset $C \subset X$. We denote $M_i = \max_{x \in C} D_i\varphi(x)$ and $M_j = \max_{x \in C} D_j\varphi(x)$. Then the change of φ between two points $x, y \in C$ must necessarily be weakly smaller than $\max(M_i, M_j)\|x - y\|$, and thus we can set $\lambda = \max(M_i, M_j)$ for every compact set $C \subset X$ such that φ is Lipschitz continuous. Hence assumption 2 applies.

The set $C \subset X$ in condition 3 denotes the state space of all *relevant* states. In our cases this is $C = [0, 1]^2$ as in the context of our model the two state vector coordinates denote shares - the shares of T players in each type population. The induced solution mapping $\xi(., x^0)$ is the solution to the dynamic problem given by the difference equation $z_T^{i,t+1} - z_T^{i,t} = \varphi_i(z_T^t)$. We have shown in section 6.3 that the population can never move past an edge (including corners) and out of C, thus every population state $\xi(t, x^0)$ with $x^0 \in C$ and $t \geq 0$ must lie within C and so assumptions 3 applies as well.

[20] See also section 6.1 in Weibull (1997).
[21] For example presented as Theorem 4.4.3 in Cunningham (2021).

Thus all three assumptions apply and we can use the result of Proposition 6.3.

A.4 Derivation of minimax regret to next round

We first derive maximum regret to the next round for all homogeneously dominant rules. Recall that regret to next round is given by the following expression if we assume it is positive:

$$r_i^{t,t+1} = -(z_T^{i,t+1} - z_T^{i,t})(\chi_i - \mu_i)$$

After inserting the population dynamics (5):

$$\begin{aligned} r_i^{t,t+1} &= -(z_T^{i,t+1} - z_T^{i,t})(\chi_i - \mu_i) = \\ &= -\Big[\lambda_i z_T^{i,t}(1 - z_T^{i,t})[s_T^i - s_B^i] + \\ &\quad + (1-\lambda_i)[(1 - z_T^{i,t})z_T^{j,t}s_T^{ij} - z_T^{i,t}(1 - z_T^{j,t})s_B^{ij}]\Big](\chi_i - \mu_i) \end{aligned}$$

We have already proven in section 4.1 that the cross-derivative is positive for all decision rules, i.e. $\partial(z_T^{i,t+1} - z_T^{i,t})/\partial z_T^{j,t} > 0$. This implies:[22]

$$\begin{aligned} \partial r_i^{t,t+1}/\partial z_T^{j,t} &< 0 \qquad \text{for } \chi_i > \mu_i \\ \partial r_i^{t,t+1}/\partial z_T^{j,t} &> 0 \qquad \text{for } \chi_i < \mu_i \end{aligned}$$

We therefore have to make a case distinction:
If $T \succ_i B \Leftrightarrow \chi_i > \mu_i$, then regret will be maximized at $z_T^{j,t} = 0$.
If $T \prec_i B \Leftrightarrow \chi_i < \mu_i$, then regret will be maximized at $z_T^{j,t} = 1$.
Note that regret is trivially zero if both options have the same expected payoff.

Maximum regret thus becomes:

$$\max r_i^{t,t+1} = \max\Big[\max_{\substack{z_T^{i,t},\lambda_i,\\ \chi_i,\mu_i,\chi_j,\mu_j\\ \chi_i>\mu_i}} r_i^{t,t+1}(z_T^{j,t}=0),\ \max_{\substack{z_T^{i,t},\lambda_i,\\ \chi_i,\mu_i,\chi_j,\mu_j\\ \chi_i<\mu_i}} r_i^{t,t+1}(z_T^{j,t}=1)\Big]$$

Due to symmetry between the two alternatives we will find equivalent expressions for either case if we continue the maximization. In particular, the

[22] For the sake of completeness we note that this derivative will be zero if both options have the same expected utility as then regret is constant at zero.

above maximum values in either case will be equal. Thus we only consider the case $T \succ_i B$ in the following. Note that this implies $\chi_i > \mu_i$ and $s_T^i > s_B^i$. So we find:

$$\max r_i^{t,t+1} = \max_{\substack{z_T^{i,t},\lambda_i,\\ \chi_i,\mu_i,\chi_j,\mu_j\\ \chi_i>\mu_i}} r_i^{t,t+1}(z_T^{j,t} = 0)$$

We continue:

$$\begin{aligned} r_i^{t,t+1}(z_T^{j,t} = 0) &= -\left[\lambda_i z_T^{i,t}(1 - z_T^{i,t})[s_T^i - s_B^i] - (1-\lambda_i) z_T^{i,t} s_B^{ij}\right](\chi_i - \mu_i) = \\ &= z_T^{i,t}\left[(1-\lambda_i)s_B^{ij} - \lambda_i(1 - z_T^{i,t})[s_T^i - s_B^i]\right](\chi_i - \mu_i) \end{aligned}$$

We see that regret will only be positive and non-zero if the expression in brackets is positive (as $\chi_i > \mu_i$):

$$(1-\lambda_i)s_B^{ij} - \lambda_i(1 - z_T^{i,t})[s_T^i - s_B^i] > 0 \Leftrightarrow z_T^{i,t} > \frac{\lambda_i(s_T^i - s_B^i) - (1-\lambda_i)s_B^{ij}}{\lambda_i(s_T^i - s_B^i)}$$

If this condition is not met then regret will simply be zero. Note that after the last equivalence the denominator on the RHS is positive and that the whole RHS will always be smaller or equal to 1 and can potentially be negative. Further recall that we require $z_T^{i,t} \geq 0$. The condition for regret to be positive is thus:

$$z_T^{i,t} > \max\left[0, \frac{\lambda_i(s_T^i - s_B^i) - (1-\lambda_i)s_B^{ij}}{\lambda_i(s_T^i - s_B^i)}\right] \tag{15}$$

We now form the derivative to continue with regret maximization:

$$\begin{aligned} \partial r_i^{t,t+1}(z_T^{j,t} = 0)/\partial z_T^{i,t} &= \left[(1-\lambda_i)s_B^{ij} - \lambda_i(1 - z_T^{i,t})[s_T^i - s_B^i]\right](\chi_i - \mu_i) + \\ &+ z_T^{i,t}\left[\lambda_i(s_T^i - s_B^i)\right](\chi_i - \mu_i) = \\ &= \left[(1-\lambda_i)s_B^{ij} - \lambda_i(1 - 2z_T^{i,t})[s_T^i - s_B^i]\right](\chi_i - \mu_i) \end{aligned}$$

The second derivative is:

$$\partial^2 r_i^{t,t+1}(z_T^{j,t} = 0)/(\partial z_T^{i,t})^2 = 2\lambda_i(s_T^i - s_B^i)(\chi_i - \mu_i)$$

The second derivative is always positive as $s_T^i - s_B^i > 0$ and $\chi_i - \mu_i > 0$. The value of $z_T^{i,t}$ that maximizes regret therefore depends on where the minimum is located. The minimum is attained at:

$$\begin{aligned}
\partial r_i^{t,t+1}(z_T^{j,t}=0)/\partial z_T^{i,t} &= 0 \\
\left[(1-\lambda_i)s_B^{ij} - \lambda_i(1-z_T^{i,t})[s_T^i - s_B^i]\right](\chi_i - \mu_i) &= 0 \\
2\lambda_i z_T^{i,t}(s_T^i - s_B^i) &= \lambda_i(s_T^i - s_B^i) - (1-\lambda_i)s_B^{ij} \\
z_{T,\min}^{i,t} &= \frac{\lambda_i(s_T^i - s_B^i) - (1-\lambda_i)s_B^{ij}}{2\lambda_i(s_T^i - s_B^i)}
\end{aligned}$$

We have shown that regret is positive if condition (15) holds. We see that this condition is fulfilled only to the right of the above minimum value. As the second derivative is always positive this implies that maximum regret is attained at the maximum possible value, i.e. at $z_T^{i,t} = 1$:

$$\max r_i^{t,t+1} = \max_{\substack{\lambda_i,\\ \chi_i,\mu_i,\chi_j,\mu_j \\ \chi_i > \mu_i}} (1-\lambda_i)s_B^{ij}(\chi_i - \mu_i)$$

We can now maximize with regard to the type share λ_i. Here we immediately see that it should be as small as possible for regret to be maximized. However, it cannot be equal to zero as then type i would not exist in the population. That being said, we still assume that it approaches zero and find for maximum regret[23]:

$$\max r_i^{t,t+1} = \max_{\substack{\chi_i,\mu_i,\chi_j,\mu_j \\ \chi_i > \mu_i}} s_B^{ij}(\chi_i - \mu_i)$$

Note that s_B^{ij} is in general a function of the valuation parameters $(\chi_i, \mu_i, \chi_j, \mu_j)$ as we know from (4):

$$s_B^{ij} = \chi_i\mu_j a + (1-\chi_i)\mu_j b + \chi_i(1-\mu_j)c + (1-\chi_i)(1-\mu_j)d$$

As we limit our attention to homogeneously dominant rules it holds that $b = 1$ and $c = 0$. Thus:

$$s_B^{ij} = \chi_i\mu_j a + (1-\chi_i)\mu_j + (1-\chi_i)(1-\mu_j)d$$

Then maximum regret to next round becomes:

$$\max r_i^{t,t+1} = \max_{\substack{\chi_i,\mu_i,\chi_j,\mu_j \\ \chi_i > \mu_i}} \left[\chi_i\mu_j a + (1-\chi_i)\mu_j + (1-\chi_i)(1-\mu_j)d\right](\chi_i - \mu_i)$$

[23]Strictly speaking this is the supremum of regret to the next round.

We immediately see that the value of χ_j does not matter as it is missing from this expression. This makes sense, as we have found that $z_T^{j,t} = 0$ and thus no player of the other type is playing T. Hence regret is maximized for $\chi_j \in [0, 1]$.
We further see that regret is maximized for $\mu_i = 0$.

Then maximum regret to next round becomes:

$$\max r_i^{t,t+1} = \max_{\substack{\chi_i,\mu_j \\ \chi_i>0}} \left[\chi_i \mu_j a + (1-\chi_i)\mu_j + (1-\chi_i)(1-\mu_j)d\right] \chi_i$$

We define:

$$\tilde{r}_i^{t,t+1}(\chi_i, \mu_j) := \left[\chi_i \mu_j a + (1-\chi_i)\mu_j + (1-\chi_i)(1-\mu_j)d\right] \chi_i$$

Differentiation after μ_j gives us:

$$\partial \tilde{r}_i^{t,t+1}(\chi_i, \mu_j)/\partial \mu_j = \left[\chi_i a + (1-\chi_i) - (1-\chi_i)d\right] \chi_i$$

It is impossible for this expression to turn negative as $d \leq 1$. If the expression is zero, then all values of μ_j maximize regret as μ_j drops out of the original function. That is the case if $a = 0$ and $d = 1$.
For all other values of a and d the expression is positive for all $\mu_j \in [0, 1]$ and thus $\mu_j = 1$ maximizes regret.

In the first case we find:

$$\tilde{r}_i^{t,t+1}(\chi_i; a = 0, d = 1) = (1-\chi_i)\chi_i$$

This expression is maximized for $\chi_i = 0.5$ with a maximum regret of $\max r_i^{t,t+1} = 0.25$. We compare this to the second case.

In the second case we have:

$$\max r_i^{t,t+1} = \max_{\substack{\chi_i \\ \chi_i>0}} \left[\chi_i a + (1-\chi_i)\right] \chi_i = \max_{\substack{\chi_i \\ \chi_i>0}} \left[1-(1-a)\chi_i\right] \chi_i$$

Differentiation after χ_i gives us:

$$\partial \tilde{r}_i^{t,t+1}(\chi_i; \mu_j = 1)/\partial \chi_i = \left[1-(1-a)\chi_i\right] - (1-a)\chi_i = 1 - 2(1-a)\chi_i \overset{!}{=} 0$$

Regret is maximized here for $\chi_i = \frac{1}{2(1-a)}$ with a maximum regret of $\max r_i^{t,t+1} = \frac{1}{4(1-a)}$. Thus we can summarize:

$$\max r_i^{t,t+1} = \begin{cases} 0.25 & \text{if } a = 0, d = 1 \\ \frac{1}{4(1-a)} & \text{else} \end{cases}$$

If we now minimize maximum regret we get for the second case $a = 0$, $d \in [0, 1)$ with a maximum regret of $\max r_i^{t,t+1} = 0.25$, the same value as in the first case.

Recall that ρ - PIR is the homogeneously dominant rule with $(a, d) = (0, 0)$ and PRR the one with $(a, d) = (0, 1)$. Thus we have now found that maximum regret to next round is minimized by ρ - PIR and PRR as well as all convex combinations of them.