Introspective equilibrium with stereotypes in diverse groups

A theoretical approach to stereotypes influence in coordination games

Author: Lukas Bartuseck
October 2021
Master essay 2 (NEKP01)
Supervisor: Roel van Veldhuizen, Lund University

Abstract

In this paper, I show how stereotypes can confound the effects cultural diversity has on economic outcomes. This is done by expanding Kets and Sandroni's (2021) theory of introspective equilibrium and cultural diversity by allowing inaccurate simplified beliefs about cultural groups. These simplified views (stereotypes) affect how strategic uncertainty is affected by diversity, the key mechanism of Kets and Sandroni's model. This extension allows for more realism when modeling diverse groups while keeping the original model's results.

Keywords: introspective equilibrium, stereotypes, diversity, coordination game,

Table of content

1. Introduction

Stereotypes are a common topic, both in society as well as in research. With increasing diverse societies, understanding stereotypes also becomes more important. Since stereotypes affect how people perceive groups different from them, it is interesting to research how this might affect economic interaction. For example, stereotypes could cause miscoordination or inefficient equilibria, leading to welfare losses. Therefore, it is relevant to better understand the effect of stereotypes in economic interaction.

To do this my goal is to create a model of strategic interaction in which stereotypes affect the perception of players. This would create a theoretical framework to see how perception through stereotypes influences the coordination on an equilibrium. As my focus lies on the perception of others I abstract from potential effects on payoffs when people have stereotypes.

One model that is very close to this idea is Kets and Sandroni's (2021) model about introspective equilibrium. There the authors incorporate culture as a cognitive factor into level-k theory to model how in a coordination game groups that are identical in payoff relevant factors end up at different equilibria. In essence, agents belonging to one of two culturally distinct groups play a coordination game with multiple equilibria. Starting with an instinctive reaction (impulse) to the game, players reflect on their impulse to estimate the other players' action and best respond in accordance. Reflecting further, agents recognize that other players will have done the same and formulate a best response to other players' best response. This iterative process continues until all players stop altering their choice of action. The resulting strategy profile is called the introspective equilibrium. In this process, a player's culture shapes the way they build estimates of other players.

Using this relationship between a player's perception and their culture, I expand the model by allowing players to hold constant simplified beliefs about players from another group adding a bias to their beliefs. This way I add stereotypes to the model and show how stereotype beliefs influence players equilibrium selection process. Based on Kets and Sandroni's key result that diversity through its effect on uncertainty can increase or decrease welfare, I demonstrate that stereotypes in this model can have the same negative or even positive effect through the same mechanism.

This paper is organized as follows. Chapter 2 explains Kets and Sandroni's (2021) original model and the coordination game that I focus on, and Chapter 3 introduces the concept of stereotypes to said model. In chapter 4 I show that an introspective equilibrium still exists after adding negative stereotypes, how it affects uncertainty and in turn welfare. Chapter 5 concludes this paper.

2. Kets and Sandroni's model and results

My model focuses on the concept of introspective equilibrium from Kets and Sandroni (2021) and how to add the idea of stereotypes to it. For this, I will start introducing their work on introspective equilibrium first since my model builds on theirs. The main research objective of Kets and Sandroni (2021) is to look at how societies that differ in cultural aspects but are equal in the payoff-dimension end up in different equilibria. To replicate these situations they use different types of coordination games and explicitly model players' reasoning process about others. They approach this by creating a model in which culture shapes players' perception and analyze how cultural diversity, measured by the size of the minority, can explain the evidence that diversity is a "double-edged sword" (Swidler, 1986 and Millikens and Martins, 1996).

2.1. The game

While Kets and Sandroni (2021) analyze multiple coordination games with strategic complementarities, this paper focuses on their linear game with identical preferences. The game at hand is a coordination game with two actions $s_j \in \{H, L\}$, High (H) and Low (L), and with a continuum of (risk-neutral) players $N = [0, 1]$. There are two Nash-equilibria, everyone choosing H or everyone choosing action L with the "High" equilibrium having a larger payoff than the "Low" equilibrium. It is however unclear which equilibrium players coordinate on, creating strategic uncertainty[1]. The situation of an individual player can be visualized as

	m
H	$u(H, m)$
L	$u(L, m)$

With m being the proportion of players choosing action H.

[1] Kets and Sandroni (2021) state that "traditional game theory" disposes of the uncertainty by assuming players select one of the equilibria, which makes it unfitting for the research objective.

As already mentioned, the game has strategic complementarities which means that the more players choose the same action the higher the incentive to join them. In addition, as it is a linear game, it follows that $u(H, m) - u(L, m)$ increases with m in a linear fashion. Next, Kets and Sandroni (2021) standardize payoffs for H=1 and L=0. The resulting payoff function of player j with action s_j is given by

$$-[(s_j - m)^2 + (s_j - \tau)^2]$$

$\tau \in [0, 1]$ can be interpreted as the economic environment of the game, indicating how strong the potential payoffs of action H differs from that of action L. This defines the risk factor of the game $\rho = 1 - \tau$, showing that the more the payoffs differ the lower the risk associated with choosing action. H. If $\tau = 1$, payoff of action H would at least be as high as that of action L, independent of m. Intuitively, there would be no risk in choosing H. From the payoff function Kets and Sandroni (2021) show that $u(H, m) = u(L, m)$ iff $m = 1 - \tau = \rho$. Thus, it is a player's best response to choose action H if $m \geq \rho$. With players not knowing which action is taken by everyone else, there is uncertainty about which action is optimal.

2.2. Introspective process

In Kets and Sandroni's (2021) model players resolve this uncertainty by taking another person's perspective. This is called introspective process and is based on the theory of mind, a concept from psychology (Apperly (2012)). This goes as follows:

At the start of the game all players receive a private signal I (called impulse) from a commonly known distribution. Their payoffs are unaffected by the signal. The set of signals equals the set of actions $\{H, L\}$ and receiving a signal means having an instinctive reaction to choose that action. Acting on the impulse is called the level-0 strategy. However, players proceed by reflecting on their impulse, realizing that everyone else also has an impulse and formulate a best response to their impulses (level-1). Reflecting further, they realize that other players will do the same as they did at level-1 and adjust their best response if necessary. This will continue until no player desires to alter their action. This can be formalized as:

Players…

(i) choose the action indicated by their impulse (level-0)

(ii) calculate the best response assuming that all other players follow their impulses (level-1),

(iii) calculate the best response assuming that all others behave as described by (ii) (level-2),

(iv) calculate the best response assuming that all others behave as described by (iii) (level-3), …

The limit of this process (level-$k \to \infty$) is the introspective equilibrium. Note that the impulse does not affect the payoff function. The underlying process is the same as the level-k theory (iterative reasoning). What is new is the impulse governing level-0 behavior, which allows Kets and Sandroni (2021) to interpret the resulting equilibrium as one influenced by culture.

2.3. Cultural diversity

Next is Kets and Sandroni's assumption about how culture influences players' impulses and perception about them, which is based on research from sociology (DiMaggio, 1997) and anthropology (D'Andrade, 1995). To introduce the concept of culture into their model Kets and Sandroni put players in two groups $G \in \{A, B\}$ with weights α for group A and β for group B. Players belong to one of those groups ($\alpha + \beta = 1$), group membership is observable and does not affect payoffs. Group A is defined as the majority and B as the minority ($\beta \leq 0.5$). The higher β the more diverse the population. As explained before, every player receives an impulse at the beginning of the game. To abstract from systematic differences between action H and L on an impulse level, Kets and Sandroni (2021) set the prior probability of impulse H as $p = 0.5$. This means that players are not instinctively drawn to one action more than the other. What draws players to an action are only payoff salience (τ) and their culture as I will explain now. When players reflect and calculate the best response according to (ii) they build posterior beliefs about other players' impulses. What players expect from their own group and others depends on how similar they are, which is determined by a groups culture. For their own group this is given by the variable $q \in (0.5, 1)$, which stands for "culture strength" of a group. The stronger the culture the more similar are their members and you expect more players to share your impulse. This means that when a player receives an impulse, of her group she expects a proportion of $Q_{in} \coloneqq q^2 + (1-q)^2$ to have the same impulse as her. From there, a player builds the posterior belief about people from another group sharing her impulse by how similar their cultures are. Kets and Sandroni (2021) measure this through the variable $d \in (0, 1)$, standing for cultural distance. As the name suggests, a high value of d means that cultures are unsimilar. This results in the expectation $Q_{out} \coloneqq 0.5d + (1-d)Q_{in}$ with $Q_{in} > Q_{out} > 0.5$. This reflects that players see members of their own group as more similar to themselves and thus expect a higher share of them to have the same impulse. Consider that players with impulse H and players with impulse L both expect the same proportion of players to share their impulse. In total, a majority player expects $Q^{maj} = \alpha Q_{in} + \beta Q_{out}$ to share his impulse while for a minority player it is $Q^{min} = \alpha Q_{out} + \beta Q_{in}$.

2.4. Kets and Sandroni's theoretical results

With this it is possible to calculate players expectations about m. Player's level-1 conditional expectation about the proportion of players choosing action H is given by

$$\mathbb{E}^1[m \mid I_j = H, G_j = A] = \alpha Q_{in} + \beta Q_{out} \coloneqq \rho^1_{HA}$$
$$\mathbb{E}^1[m \mid I_j = H, G_j = B] = \beta Q_{in} + \alpha Q_{out} \coloneqq \rho^1_{HB}$$
$$\mathbb{E}^1[m \mid I_j = L, G_j = B] = \beta \tilde{Q}_{in} + \alpha \tilde{Q}_{out} \coloneqq \rho^1_{LB}$$
$$\mathbb{E}^1[m \mid I_j = L, G_j = A] = \alpha \tilde{Q}_{in} + \beta \tilde{Q}_{out} \coloneqq \rho^1_{LA}$$
$$with\ \tilde{Q} = 1 - Q$$

Remember that for a player to choose H as a best response he needs to expect at least $m = \rho$. Thus, the expectation about m can be interpreted as the maximum risk under which a player chooses High. So if $\rho < \rho^k_{IG}$ it means that the risk of the game is sufficiently low for a player of group G with impulse I to choose action H.

Lastly, the uncertainty a player faces is given by the variance of impulses he expects. It follows that the uncertainty given by variable V is

$$V^{maj} = Q^{maj} = Q^{maj}(1 - Q^{maj})$$
$$V^{min} = Q^{min} = Q^{min}(1 - Q^{min})$$

The aggregate uncertainty is given by their weighted average

$$Var = \alpha Var^{maj} + \beta Var^{min}$$

From $Q_{in} > Q_{out} > 0.5$ follows that a minority player faces greater strategic uncertainty. Based on this, Kets and Sandroni's (2021) show that by increasing diversity through β, the resulting increase in uncertainty decreases or increases the probability of inefficient lock-ins and miscoordination, depending on the economic environment (τ). Essentially, if players have only their culture or only the games payoff as orientation it is clear which action will be picked, and there is going to be little miscoordination. Increasing diversity and thus uncertainty in the first situation would increase miscoordination and as a result reduce welfare. If both equilibria stand out, "Low" in terms of culture, "High" in terms of payoff, players might go for the "Low" equilibrium even though they get lower payoffs because they believe others will be culturally inclined to do the same[2]. Increasing diversity (higher strategic uncertainty) will weaken the payoff sub-optimal equilibrium and drive players towards the efficient equilibrium. With this Kets and Sandroni's (2021) theory can explain some of the disparate evidence about cultural diversity (Swidler, 1986 and Millikens and Martins, 1996).

3. Stereotypes

Standard theory and the model so far assume that people update beliefs in a rational way. In reality, people may not have accurate beliefs because of stereotypes. In particular, stereotypes may distort beliefs in a way that people underestimate or overestimate the prevalence of impulses in some part of the population. To model this I introduce stereotypes into the introspective equilibrium process of Kets and Sandroni (2021) and analyze the effects on it. In order to introduce stereotypes, it is necessary to add certain assumptions. First, assume that some players have stereotypes. This is a very plain assumption that introduces the idea of stereotypes to the preexisting model. As stereotypes are fairly common as well as research about them (Dovidio et al., 1985; Bender Peterson, 1990; Ellemers, 2018), this is reasonable to assume. Second, let stereotypes be a given belief about the distribution of impulses in a cultural group, shared by a part of the population. This assumption shapes how stereotypes behave in the model and is based on the non-academic and academic definition of stereotypes. According to the Cambridge Dictionary, a stereotype is "to have a set idea about what a particular type of person is like, especially an idea that is wrong". However, in academia the definition is not that clear-cut as it is pointed out by Kanahara (2006). The author identifies beliefs and group concepts as the two most important components; thus it seems intuitive that in this model stereotypes affect the beliefs about impulse distribution. It is important to note that having stereotypes does not affect the payoff function. Here, stereotypes do not mean a distaste about playing with the stereotyped. In that aspect it is closer to the concept of statistical discrimination rather than taste-based discrimination. Furthermore, the assumption simplifies stereotypes to a binary: either a player has a stereotype or not. These are the fundamental assumptions of my

[2] If players actually choose the low equilibrium in this situation depends on m and ρ

model as they introduce stereotypes to Kets and Sandroni's (2021) introspective equilibrium and define how they work in it.
To analyze the effects of stereotypes I introduce some simplifying assumptions. Let there be only one stereotype at a time. While it is clearly possible that more than one stereotype is present at a time, to analyze the effect a stereotype has on the introspective equilibrium, I allow for only one stereotype at a time. Moreover, a stereotype is about a culturally different group, which is to exclude self-stereotyping from the model.
Lastly, given the scope of this paper, I will only look at stereotypes about the minority group from here on. The general effect a stereotype has on the introspective equilibrium should not depend on which side holds stereotypes, thus reducing it to one group should not affect the general result of the model.

The next step is to translate the assumptions above into the actual model. To introduce stereotypes (assumption 1) I split group A into a part with stereotypes and one without stereotypes. Let $\gamma \in (0,1)$ be the proportion of majority players holding stereotypes, then $\gamma\alpha$ is the proportion of all players that have stereotypes. To address players with and without stereotypes separately, let majority players without stereotypes belong to group A (as before) and the new type of players belong to a new group, group C.
For groups A and C to work as one cultural group I define $d_{AC} = 0$. That way both groups treat each other as one ($d = 0 \Leftrightarrow Q_{in} = Q_{out}$ between group A and C) and $d_{AB} = d_{CB}$, group B expects the same impulse distribution from group C as they do from group A. Note that the distribution of impulses is unchanged, what changes is the distribution of beliefs about impulses, as I will explain next.

Regarding assumption 2, it defines a stereotype as a given belief about the distribution of impulses. To reflect this in the model I add a new variable $c \in [0, 1]$. c is the proportion of minority players expected by group C to have impulse H. To clarify, players with stereotypes have the same expectation about players from the minority, independent of the impulse they received. For those with stereotypes, $Q_{out} = 0.5d + (1-d)Q_{in}$ gets replaced by the stereotype belief c. As Q indicates how many players share your impulse, for players of group C it depends on the impulse they received. That is $Q_{out}^{HC} = c$, $Q_{out}^{LC} = 1 - c$ for stereotype players with impulse H and L respectively. Modeling how c is formed is outside the scope of this paper hence it stays an exogenous variable throughout this paper. I also add another constraint on stereotypes here. That is I'm limiting stereotypes to values of $c < 1 - Q_{out}$ and $c > Q_{out}$. This way the bias a stereotype causes is either negative or positive for both high impulse and low impulse players. For values in between, the same stereotype causes a negative bias for one part of the group while simultaneously causing a positive bias for the other part of the same group which seems counterintuitive to the idea of a stereotype. $c < 1 - Q_{out}$ can be interpreted as a negative stereotype, $c > Q_{out}$ as a "positive" stereotype.

What does this mean for the introspective process? How the introspective process works in general is unaffected. At level-0 players follow their impulse, at level-1 players best respond to level-0 based on their expectations about impulse distribution, at level-2 players best respond to the choices at level-1 which in turn are based on different expectations regarding impulse distribution. This is the same as a model without stereotypes. What is different is $\rho_{HC}^1 \neq \rho_{HA}^1$ and $\rho_{LC}^1 \neq \rho_{LA}^1$ (the maximum risk a player is willing to accept) through the change in Q_{out} for group C. Depending on the extent of change and the economic environment, the stereotype belief can cause a shift in the action chosen. At level-2 the choices influenced by stereotypes are considered by every player and as a consequence, the stereotypes of some indirectly affect the behavior of those without as well. However, this does not mean that players drop their stereotype or others adopt them. To put it simply, they expect others to react, stereotype or not, and best respond accordingly.

Besides how they build expectations about the minority group, they behave the same as the other groups. This means everyone goes through the same introspective process as explained in chapter 2 and they still have the same payoff function. One might wonder how stereotype beliefs survive the introspective process. While the process requires to take other people's perspective to best respond, it does not cause players to adopt those beliefs. (On level-1 you best respond according to your own beliefs, on level-2 you best respond according to the beliefs of everyone, including your own (stereotype) beliefs…). This resistance to change reflects results found in research about the persistence of stereotypes (see Johnston (1996), Keita and Kittles (1997) and Grossman (2013) among others).

4. Theoretical Results

Before showing the effect of stereotypes on strategic uncertainty, I first demonstrate that introspective equilibria still exist after the addition of stereotypes. That is, it converges to an introspective equilibrium for $k \to \infty$. This can be confirmed by repeating the proof by Kets and Sandroni (2021) after adding the two new combinations of group and impulse. Here, I will demonstrate it for one case of a negative stereotype, $c < 1 - Q_{out}$ (The other cases for a negative can be found in the appendix)

To do this I need to calculate the level-1 conditional expectations about the share of players choosing action H (m) first. Those are:

$$\mathbb{E}^1[m \mid I_j = H, G_j = A] = \alpha(1-\gamma)Q_{in} + \beta Q_{out} + \alpha\gamma Q_{in} = \alpha Q_{in} + \beta Q_{out} \coloneqq \rho_{HA}^1$$
$$\mathbb{E}^1[m \mid I_j = H, G_j = B] = \beta Q_{in} + \alpha Q_{out} \coloneqq \rho_{HB}^1$$
$$\mathbb{E}^1[m \mid I_j = H, G_j = C] = \alpha(1-\gamma)Q_{in} + \beta c + \alpha\gamma Q_{in} = \alpha Q_{in} + \beta c \coloneqq \rho_{HC}^1$$
$$\mathbb{E}^1[m \mid I_j = L, G_j = B] = \beta \tilde{Q}_{in} + \alpha \tilde{Q}_{out} \coloneqq \rho_{LB}^1$$
$$\mathbb{E}^1[m \mid I_j = L, G_j = A] = \alpha(1-\gamma)\tilde{Q}_{in} + \beta \tilde{Q}_{out} + \alpha\gamma \tilde{Q}_{in} = \alpha \tilde{Q}_{in} + \beta \tilde{Q}_{out} \coloneqq \rho_{LA}^1$$
$$\mathbb{E}^1[m \mid I_j = L, G_j = C] = \alpha(1-\gamma)\tilde{Q}_{in} + \beta c + \alpha\gamma \tilde{Q}_{in} = \alpha \tilde{Q}_{in} + \beta c \coloneqq \rho_{LC}^1$$
$$with\ \tilde{Q} = 1 - Q$$

Note that for a player to choose action H after introspection $\rho < \rho_{IG}^k$ needs to be fulfilled. H is a best response iff $m < \rho$, (chapter 2). If it is, enough players are expected to choose H for a player to choose H themselves given the risk factor of the game (ρ). These expectations determine for what degree of risk (ρ) players choose action H. In that sense ρ_{IG}^1 is the maximum risk under which a player chooses action H at level-1.

Next, I need to determine the (ordinal) ranking of the expectations. As shown by Kets and Sandroni (2021), from $Q_{in} > Q_{out} > \frac{1}{2}$ and $\alpha > \beta$ follows

$$\rho_{HA}^1 > \rho_{HB}^1 > \rho_{LB}^1 > \rho_{LA}^1$$

For negative stereotypes, $c < 1 - Q_{out} \to \rho_{LA}^1 > \rho_{LC}^1$.

While it is clear that $\rho_{HA}^1 > \rho_{HC}^1 > \rho_{LA}^1$, the precise placement of ρ_{HC}^1 depends on c, Q_{in} and Q_{out}. What can be observed here is that, unsurprisingly, negative stereotypes reduce the scope of cooperating on High, since a negatively biased stereotype reduces expectations.

With the relative placing of ρ_{HC}^1 not being definitive there are a lot of cases to consider. To proof the existence of introspective equilibrium in general is outside the scope of this paper. To continue I will assume the case of

$$\rho_{HA}^1 > \rho_{HB}^1 > \rho_{HC}^1 > \rho_{LB}^1 > \rho_{LA}^1 > \rho_{LC}^1$$

With the strategy profile $\sigma^k = (\sigma_{HA}^k, \sigma_{HB}^k, \sigma_{HC}^k, \sigma_{LB}^k, \sigma_{LA}^k, \sigma_{LC}^k)$ at level-0 being $\sigma^0 = (H, H, H, L, L, L)$ it creates 7 cases to consider. Those are

(1) $\rho < \rho_{LC}^1 \Rightarrow \sigma^1 = (H, H, H, H, H, H)$
(2) $\rho \in (\rho_{LC}^1, \rho_{LA}^1) \Rightarrow \sigma^1 = (H, H, H, H, H, L)$
(3) $\rho \in (\rho_{LA}^1, \rho_{LB}^1) \Rightarrow \sigma^1 = (H, H, H, H, L, L)$
(4) $\rho \in (\rho_{LB}^1, \rho_{HC}^1) \Rightarrow \sigma^1 = (H, H, H, L, L, L)$
(5) $\rho \in (\rho_{HC}^1, \rho_{HB}^1) \Rightarrow \sigma^1 = (H, H, L, L, L, L)$
(6) $\rho \in (\rho_{HB}^1, \rho_{HA}^1) \Rightarrow \sigma^1 = (H, L, L, L, L, L)$
(7) $\rho > \rho_{HA}^1 \Rightarrow \sigma^1 = (L, L, L, L, L, L)$

Now it is to determine the introspective equilibria for the different cases. For ease of reference I'm going to refer to players by their group and impulse, that means minority players with high impulse are HB and stereotype players with low impulse are LC. Starting with the easiest cases, (1) is an introspective equilibrium already. As at level-1 everyone best response is action H it clearly follows that at level-2 and higher H is also the best response. This is because for all $k > 1, \mathbb{E}^k[m] = 1 > \rho_{IG}$ for all I, G.
Case (7) is similar. With everyone playing action L at level-1 the expected share to choose action H is zero and thus at level-2 everyone still chooses L (formally, for all $k > 1, \mathbb{E}^k[m] = 0 < \rho_{IG}$ for all I, G). The same can be observed for $k > 2 \Rightarrow$ In introspective equilibrium all players choose L.
Case (4) is an introspective equilibrium where everyone picks the action indicated by the impulse at any level k. At level-2 and onwards the proportions expected to choose H are the same as at level-1 resulting in equilibrium. This is because at level-1 players best response is to pick the action indicated by their impulse, so at level-1 everyone does what they would have done at level-0. Consequently, the best response at level-2 is the same as at level-1, since at level-2 everyone expects others to do what they would have done at level-0.

The remaining cases require higher levels of introspection.

For case (2) $\rho \in (\rho^1_{LC}, \rho^1_{LA})$, everyone that chose H at level-1 chooses H at higher levels as well. This is because of the strategic complementarities mentioned when I introduced the game (the more players choose H the higher the incentive to do the same). At level-1 more players choose action H as expected (from level-0 to level-1 LB and LA switch from L to H) making action H more attractive. Hence, those that choose action H at level-1 (with lower expectations) continue to do so at higher levels. LC Players remain to be considered in case (2). At level-2 they realize that low impulse players of group A and B will switch actions. Therefore, the incentive changes to $\mathbb{E}^2[m \mid I_j = L, G_j = C] = 1 - \alpha\gamma Q_{in} \coloneqq \rho^2_{LC}$. If $\rho < \rho^2_{LC}$, then all players choose H at level-$k \geq 2 \Rightarrow$ in introspective equilibrium everyone chooses H. If $\rho > \rho^2_{LC}$, then LC players choose L at level-$k \geq 2$. With everyone but LC choosing action H the expected value of m remains the same for level-2 and onward. Therefore on higher levels LC players will continue to choose L and in introspective equilibrium everyone but LC plays action H.

Case (3): LA and LC choose action L, the rest chooses H. Those that picked H at level-1 will do so at higher levels by strategic complementarities just like above which leaves LA and LC. At level-2 LA will choose H if: $\mathbb{E}^2[m \mid I_j = L, G_j = A] = 1 - \alpha Q_{in} \coloneqq \rho^2_{LA} > \rho$. The same goes for LC with $\mathbb{E}^2[m \mid I_j = L, G_j = C] = 1 - \alpha Q_{in} \coloneqq \rho^2_{LC} = \rho^2_{LA}$. With the same reason as before, if $\rho < \rho^2_{LA}$ then in introspective equilibrium everyone chooses H, if $\rho > \rho^2_{LA}$ then low impulse players from the majority group choose L while the rest chooses H in introspective equilibrium.

Case (5): HA and HB choose H, the rest chooses L. By strategic complementarities, those that choose L at level-1 continue to do so at higher levels. For HA, $\mathbb{E}^2[m \mid I_j = H, G_j = A] = (1-\gamma)\alpha Q_{in} + \beta Q_{out} \coloneqq \rho^2_{HA}$ and for HB, $\mathbb{E}^2[m \mid I_j = H, G_j = B] = (1-\gamma)\alpha Q_{out} + \beta Q_{in} \coloneqq \rho^2_{HB}$.

$\rho^2_{HA} \geq \rho^2_{HB}$ and $\rho^2_{HA} \leq \rho^2_{HB}$ is possible, so there are four cases to consider here.

First, $\rho < \min\{\rho^2_{HA}, \rho^2_{HB}\} \Rightarrow$ HA and HB choose H, hence in introspective equilibrium high impulse players without stereotypes choose H, while the rest chooses L.

Second, $\rho > \max\{\rho^2_{HA}, \rho^2_{HB}\} \Rightarrow$ HA and HB choose L, in introspective equilibrium everyone chooses L.

Third, $\rho \in (\rho^2_{HA}, \rho^2_{HB}) \Rightarrow$ HB chooses H at level-2 and HA chooses L at level-$k \geq 2$. At level-3 HB decides based on $\mathbb{E}^3[m \mid I_j = H, G_j = B] = \beta Q_{in} \coloneqq \rho^3_{HB}$. For $\rho > \rho^3_{HB} \Rightarrow$ in introspective equilibrium everyone chooses L. For $\rho < \rho^3_{HB} \Rightarrow$ in introspective equilibrium HB chooses H, everyone else chooses L.

Fourth, $\rho \in (\rho_{HB}^2, \rho_{HA}^2)$ is mirror image of the third. HA chooses H at level-2 and HB chooses L at level-$k \geq 2$. At level-3 HA decides based on $\mathbb{E}^3[m \mid I_j = H, G_j = A] = (1-\gamma)\alpha Q_{in} \coloneqq \rho_{HA}^3$. For $\rho_{HA}^3 < \rho \Rightarrow$ in introspective equilibrium everyone chooses L. For $\rho < \rho_{HA}^3 \Rightarrow$ in introspective equilibrium HA chooses H, everyone else chooses L.

The last case to check is (6). Here, everyone but HA is choosing L at level-2 and by strategic complementarities do so at higher levels as well. This leaves HA to be considered at level-2. Same as before, $\mathbb{E}^2[m \mid I_j = H, G_j = A] = (1-\gamma)\alpha Q_{in} \coloneqq \rho_{HA}^2$ determines what HA chooses. If $\rho > \rho_{HA}^2$, everyone chooses L in introspective equilibrium. If $\rho < \rho_{HA}^2$ then in introspective equilibrium HA chooses H while the rest chooses L.

This shows that for all cases the introspective process has a limit. Bounded rationality is not required, the same as in Kets and Sandroni's (2021) theory.

For the case of positive stereotypes ($c > Q_{out}$) it follows that $\rho_{HC}^1 > \rho_{HA}^1$ and for ρ_{LC}^1 the situation is as above. In the end this creates 7 possible cases like above, showing the existence of the introspective equilibrium the same way as well. What changes is that players from group C swap their relative position regarding ρ_{IG}^1 so instead of $\sigma^0 = (\sigma_{HA}^k, \sigma_{HB}^k, \sigma_{HC}^k, \sigma_{LB}^k, \sigma_{LA}^k, \sigma_{LC}^k) = (H, H, H, L, L, L)$ it is going to be $\sigma^0 = (\sigma_{HC}^k, \sigma_{HA}^k, \sigma_{HB}^k, \sigma_{LC}^k, \sigma_{LB}^k, \sigma_{LA}^k) = (H, H, H, L, L, L)$. Thus, the introspective equilibrium should also exist for this case.

Another point that is demonstrated is that stereotypes of some players affect the behavior of others as by best responding to σ^1 they "respect" those beliefs, as mentioned at the end of last chapter. The same thing also shows that players are aware of each other's beliefs since at level-2 they best respond to every type's level-1 action. Hence there is no bounded rationality regarding the awareness of other players beliefs.

Now that I have shown that the introspective equilibrium can still exist even with stereotypes, I want to address the effects on strategic uncertainty. As strategic uncertainty is the driver behind the benefits and drawbacks of diversity as argued by Kets and Sandroni (2021), it is important to analyze which effects stereotypes have on it in this model.

Kets and Sandroni (2021, online appendix) define strategic uncertainty as $V \coloneqq Q * (1-Q)$, that is the uncertainty is highest for $Q = 0.5$. This expresses, when players expect half of all players to share their impulse, they are most uncertain.

To calculate the effect on strategic uncertainty, remember:
For majority players without stereotypes: $Q^{maj} = \alpha Q_{in} + \beta Q_{out}$
Majority players with stereotypes: $Q^{HC} = \alpha Q_{in} + \beta c$, $Q^{LC} = \alpha Q_{in} + \beta(1-c)$

The question is under what circumstances do stereotypes reduce strategic uncertainty. With strategic uncertainty maximized at $Q = 0.5$, stereotypes reduce strategic uncertainty if and only if $\left|Q^{IC} - \frac{1}{2}\right| > \left|Q^{maj} - \frac{1}{2}\right|$ [3]
Since Q_{out} depends on the impulse for stereotype players, it is possible that for HC uncertainty increases while for LC uncertainty decreases, neutralizing each other. As an example, $Q_{in}, Q_{out} \approx 0.5 \Rightarrow Q^{maj} \approx 0.5$
Put $c = 0.2$, $\alpha = 0.6 \rightarrow Q^{HC} = 0.6 * 0.5 + 0.4 * 0.2 = 0.38 \rightarrow$ decrease in uncertainty
$\rightarrow Q^{LC} = 0.6 * 0.4 + 0.4 * 0.8 = 0.62 \rightarrow$ decrease in uncertainty
For $c > Q_{out}$ it works the same, mirroring the example from above.
Put $c = 0.8$, $\alpha = 0.6 \rightarrow Q^{HC} = 0.6 * 0.5 + 0.4 * 0.8 = 0.62 \rightarrow$ decrease in uncertainty
$\rightarrow Q^{LC} = 0.6 * 0.4 + 0.4 * 0.2 = 0.38 \rightarrow$ decrease in uncertainty

The examples also demonstrate that in a situation of high strategic uncertainty, stereotypes with a clear bias reduce strategic uncertainty. Since the aggregate strategic uncertainty is a weighted average of the groups' strategic uncertainty and the stereotype-group substitutes part of the majority group, a reduction in strategic uncertainty from no stereotype to with stereotype causes a reduction in aggregate strategic uncertainty.
What this means in the context of the model is that in situations of high strategic uncertainty, a negative stereotype reduces the level of uncertainty. In accordance with Kets and Sandroni's (2021) result that strategic uncertainty moderates the tradeoff between miscoordination and inefficient lock-in in coordination games, this can have a negative or positive impact on welfare. For negative stereotypes the decrease in uncertainty can reduce miscoordination and therefore have a positive impact on welfare. On the other hand it, could also facilitate inefficient lock-in as it drives players to the suboptimal equilibrium, reducing welfare in return. Positive stereotypes work the other way around. The reduction in uncertainty caused by positive stereotypes may help avoid inefficient lock-in (pushes players towards the pareto optimal equilibrium→increase in welfare) or increase miscoordination (reduction in welfare). The case

[3] The opposite is also possible; stereotypes increase uncertainty if the reverse inequality holds.

of increasing uncertainty through stereotypes only has a positive effect on welfare if it helps avoid an inefficient lock-in. Note however that this theory only models the aspect of perception. By abstracting from effects on payoffs, potential disutility's for the stereotyped are excluded, for negative (Wheeler et al. (2001), Baltes and Cort, (2010)) as well as for positive stereotypes (Cheryan and Bodenhausen (2010), Gupta et al. (2011)). This means that the effects found in my model should not be interpreted as stereotypes potentially increasing welfare in a utility sense. Through its effect on players coordination, total payoffs might increase while overall utility decreases with the costs carried by the stereotyped.

5. Conclusion

I expanded on Kets and Sandroni's (2021) theory of introspective equilibrium to model how stereotype beliefs affect the strategic interaction in a linear coordination game. Allowing for simplified beliefs about other groups, my model shows that stereotypes can increase or reduce strategic uncertainty.

Following Kets and Sandroni (2021) introspective equilibrium model that shows that strategic uncertainty moderates the tradeoff between miscoordination and inefficient lock-in in a coordination game, my model also shows that under certain circumstances stereotypes can increase overall welfare. This indicates that the role of stereotypes in economic interaction may not be as one dimensional and more complex than anticipated. As my model is an expansion of Kets and Sandroni's (2021) original work, it is especially interesting to understand how stereotypes alter the effects of diversity that they showed. Furthermore, it demonstrates that their theory allows being adjusted for other cognitive factors without having to adjust the original theory by a lot, especially ones connected to cultural diversity. A drawback of this kind of expansion is that the model becomes less tractable as with every additional variable there are exponentially more combinations to consider. One of the limitations of my model is that it abstracts from stereotypes potentially affecting the payoff functions, especially potential negative effects on the utility of the stereotyped. Another limitation comes from the definition of stereotypes in combination with the rationality assumption. This creates a somewhat mild version of a stereotype where a player who has stereotypes assumes rationality of all players even though he has simplified beliefs about some of them. One could argue for a more extreme interpretation of stereotypes where stereotype players expect the stereotyped to always follow their impulse for every level of rationality.

This already shows one possible avenue for further research. Changing the definition of how stereotypes affect beliefs, and the introspective process of players would test the robustness of my findings. Another way to continue this work would be to show the effects on the introspective equilibrium more generally and to apply them to other types of coordination games. Lastly, one could endogenize stereotypes. One starting point for this could be Bordalo et al. (2016) research for example. Another avenue could be to research the relationship between uncertainty on one side and the "strength" of the stereotype c and its prevalence γ on the other. Depending on how they are connected some combinations of stereotypes and uncertainty might not be valid. Endogenizing stereotypes this way might allow to reduce the number of potential results, increasing the precision of the model.

6. Reference list

Baltes B. and Rudolph, C.W. (2010) ‘Examining the effect of negative Turkish stereotypes on evaluative workplace outcomes in Germany’, Journal of Managerial Psychology, 25(2), pp. 148–158.

Bender Peterson, S. (1990) ‘Gender stereotypes in children’s books: Their prevalence and influence on cognitive and affective development’, Gender & Education, 2(2), p. 185-197.

Bordalo, P. et al. (2016) ‘Stereotypes’, Quarterly Journal of Economics, 131(4), pp. 1753–1794

Cheryan Sapna and Bodenhausen Galen V. (2000) ‘When Positive Stereotypes Threaten Intellectual Performance: The Psychological Hazards of “Model Minority” Status’, Psychological Science, 11(5), pp. 399–402.

D’Andrade, R. G. (1995) The development of cognitive anthropology. Cambridge Univ. Press.

DiMaggio, Paul (1997) ‘Culture and Cognition’, Annual Review of Sociology, 23, pp. 263–287.

Dovidio, J. F., Evans, N. and Tyler, R. B. (1986) ‘Racial stereotypes: The contents of their cognitive representations’, Journal of Experimental Social Psychology, 22(1), pp. 22–37.

Ellemers, N. (2018) ‘Gender Stereotypes’, Annual review of psychology, 69, pp. 275–298.

Grossman, P. J. (2013) ‘Holding Fast: The Persistence and Dominance of Gender Stereotypes’, Economic Inquiry, 51(1), pp. 747–763.

Gupta, A., Szymanski, D. M. and Leong, F. T. L. (2011) ‘The “model minority myth”: internalized racialism of positive stereotypes as correlates of psychological distress, and attitudes toward help seeking’, Asian American Journal of Psychology, 2(2), pp. 101–114.

Johnston, L. (1996) ‘Resisting change: information-seeking and stereotype change’, European Journal of Social Psychology, 26(5), pp. 799–825.

Kanahara, S. (2006) ‘A Review of the Definitions of Stereotype and a Proposal for a Progressional Model’, Individual Differences Research, 4(5), pp. 306–321.

Keita, S. O. Y. and Kittles, R. A. (1997) ‘The Persistence of Racial Thinking and the Myth of Racial Divergence’, American Anthropologist, 99(3), pp. 534–544.

Kets, W. and Sandroni, A. (2021) ‘A Theory of Strategic Uncertainty and Cultural Diversity’, Review of Economic Studies, 88(1), pp. 287–333.

Milliken, F. J. and Martins, L. L. (1996), “Searching for Common Threads: Understanding the Multiple Effects of Diversity in Organizational Groups”, Academy of Management Review, 21, 402–433.

Swidler, A. (1986), “Culture in Action: Symbols and Strategies”, American Sociological Review, 51, 273–286.

Wheeler, S. C., Jarvis, W. B. G. and Petty, R. E. (2001) ‘Think Unto Others: The Self-Destructive Impact of Negative Racial Stereotypes’, Journal of Experimental Social Psychology, 37(2), pp. 173–180.

7. Appendix

Here, I will show for negative stereotypes that for the other possible rankings of ρ^1_{IG}, the introspective equilibrium still exists. As the steps for the proof are mostly similar, I will refer to the demonstration in chapter 4 if steps are identical.

The other possible rankings of ρ^1_{IG} are

$$\rho^1_{HA} > \rho^1_{HC} > \rho^1_{HB} > \rho^1_{LB} > \rho^1_{LA} > \rho^1_{LC} \ (N.2)$$
$$\rho^1_{HA} > \rho^1_{HB} > \rho^1_{LB} > \rho^1_{HC} > \rho^1_{LA} > \rho^1_{LC} \ (N.3)$$
$$\rho^1_{HA} > \rho^1_{HB} > \rho^1_{LB} > \rho^1_{LA} > \rho^1_{HC} > \rho^1_{LC} \ (N.4)$$

For $(N.2)\ \rho^1_{HA} > \rho^1_{HC} > \rho^1_{HB} > \rho^1_{LB} > \rho^1_{LA} > \rho^1_{LC} \to \sigma^0 = (H,H,H,L,L,L)$
There are again 7 different cases depending on ρ relative to ρ^1_{IG}

(1) $\rho < \rho^1_{LC} \Rightarrow \sigma^1 = (H,H,H,H,H,H)$
(2) $\rho \in (\rho^1_{LC}, \rho^1_{LA}) \Rightarrow \sigma^1 = (H,H,H,H,H,L)$
(3) $\rho \in (\rho^1_{LA}, \rho^1_{LB}) \Rightarrow \sigma^1 = (H,H,H,H,L,L)$
(4) $\rho \in (\rho^1_{LB}, \rho^1_{HB}) \Rightarrow \sigma^1 = (H,H,H,L,L,L)$
(5) $\rho \in (\rho^1_{HB}, \rho^1_{HC}) \Rightarrow \sigma^1 = (H,H,L,L,L,L)$
(6) $\rho \in (\rho^1_{HC}, \rho^1_{HA}) \Rightarrow \sigma^1 = (H,L,L,L,L,L)$
(7) $\rho > \rho^1_{HA} \Rightarrow \sigma^1 = (L,L,L,L,L,L)$

(1), (4) and (7) are already introspective equilibria for the same reasons as it was for (N.1) in chapter 4. Case (2) of (N.2) is identical to case (2) of (N.1). By strategic complementarities, everyone but LC chooses action H in introspective equilibrium and LC might change his action at level-2 based on $\mathbb{E}^2[m \mid I_j = L, G_j = C] = 1 - \alpha\gamma Q_{in} \coloneqq \rho^2_{LC}$. Case (3) of (N.2) works the same as case (3) of (N.1). Everyone but LC and LA chooses action H (strategic complementarities) and LA and LC might change their action at level-2 based on $\mathbb{E}^2[m \mid I_j = L, G_j = C] = 1 - \alpha Q_{in} \coloneqq \rho^2_{LC} = \rho^2_{LA}$.

Case (5): HB, LB, LA, LC choose L by strategic complementarities, HC and HA might change actions based on $\mathbb{E}^2[m \mid I_j = H, G_j = C] = 1 - \alpha Q_{in} \coloneqq \rho^2_{HC} = \rho^2_{HA}$. If $\rho > \rho^2_{HC}$, everyone plays action L in the introspective equilibrium. For $\rho < \rho^2_{HC}$, HC and HA choose action H, the rest chooses action L.

Case (6) for N(.2) is identical to case (6) for (N.1). omitted

For $(N.3)\ \rho_{HA}^1 > \rho_{HB}^1 > \rho_{LB}^1 > \rho_{HC}^1 > \rho_{LA}^1 > \rho_{LC}^1\ (N.3) \to \sigma^0 = (H, H, L, H, L, L)$

(1) $\rho < \rho_{LC}^1 \Rightarrow \sigma^1 = (H, H, H, H, H, H)$

(2) $\rho \in (\rho_{LC}^1, \rho_{LA}^1) \Rightarrow \sigma^1 = (H, H, H, H, H, L)$

(3) $\rho \in (\rho_{LA}^1, \rho_{HC}^1) \Rightarrow \sigma^1 = (H, H, H, H, L, L)$

(4) $\rho \in (\rho_{HC}^1, \rho_{LB}^1) \Rightarrow \sigma^1 = (H, H, H, L, L, L)$

(5) $\rho \in (\rho_{LB}^1, \rho_{HB}^1) \Rightarrow \sigma^1 = (H, H, L, L, L, L)$

(6) $\rho \in (\rho_{HB}^1, \rho_{HA}^1) \Rightarrow \sigma^1 = (H, L, L, L, L, L)$

(7) $\rho > \rho_{HA}^1 \Rightarrow \sigma^1 = (L, L, L, L, L, L)$

Case (1) and (7) are introspective equilibrium for the same reasons as explained before. (2) and (3) are also the same as for (N.1) and (N.2), there is an introspective equilibrium. Case (4), $\rho \in (\rho_{HC}^1, \rho_{LB}^1)$, at level-1 HA, HB and LB choose action H, the rest chooses action L. Since LB and HC switch choices at level-1, it is not as clear as before how players will decide at level 2. Everyone's level-2 conditional expectations are given by

$\mathbb{E}^2[m \mid I_j = H, G_j = A] = \beta + \alpha(1-\gamma)Q_{in} \coloneqq \rho_{HA}^2$

$\mathbb{E}^2[m \mid I_j = H, G_j = B] = \beta + \alpha(1-\gamma)Q_{out} \coloneqq \rho_{HB}^2$

$\mathbb{E}^2[m \mid I_j = H, G_j = C] = \beta + \alpha(1-\gamma)Q_{in} \coloneqq \rho_{HC}^2$

$\mathbb{E}^2[m \mid I_j = L, G_j = B] = \beta + \alpha(1-\gamma)(1-Q_{out}) \coloneqq \rho_{LB}^2$

$\mathbb{E}^2[m \mid I_j = L, G_j = A] = \beta + \alpha(1-\gamma)(1-Q_{in}) \coloneqq \rho_{LA}^2$

$\mathbb{E}^2[m \mid I_j = L, G_j = C] = \beta + \alpha(1-\gamma)(1-Q_{in}) \coloneqq \rho_{LC}^2$

As can be seen $\rho_{HC}^2 = \rho_{HA}^2$ and $\rho_{LA}^2 = \rho_{LC}^2$. It is down to four ρ_{IG} → it is the same situation as in in the original paper ⇒ an introspective equilibrium exists.

Case (5) $\rho \in (\rho_{LB}^1, \rho_{HB}^1)$. By strategic complementarities, LB, HC, LA and LC choose action L in introspective equilibrium. HA and HB remain to be considered. At level-2 they decide based on $\mathbb{E}^2[m \mid I_j = H, G_j = A] = \alpha(1-\gamma)Q_{in} + \beta Q_{out} \coloneqq \rho_{HA}^2$ and $\mathbb{E}^2[m \mid I_j = H, G_j = B] = \beta Q_{in} + \alpha(1-\gamma)Q_{out} \coloneqq \rho_{HB}^2$ respectively. It is however unclear what value is higher.

This creates 4 possibilities. First, if $\rho < min\{\rho_{HA}^2, \rho_{HB}^2\}$ ⇒ both will choose H in introspective equilibrium, the rest chooses action L. Second, if $\rho > max\{\rho_{HA}^2, \rho_{HB}^2\}$ ⇒ in introspective equilibrium everyone chooses action L. Third, if $\rho_{HA}^2 < \rho_{HB}^2$ and thus $\rho \in (\rho_{HA}^2, \rho_{HB}^2)$, HA will choose action L and HB action H at level-2. At level-3 HA chooses L by strategic complementarities $(\beta Q_{out} \coloneqq \rho_{HA}^3 < \alpha(1-\gamma)Q_{in} + \beta Q_{out} \coloneqq \rho_{HA}^2)$. HB chooses H if $\mathbb{E}^3[m \mid I_j = H, G_j = B] = \beta Q_{in} \coloneqq \rho_{HB}^3 > \rho$. If so, HB chooses H at level-$k \geq 3$, if not HB chooses L at $k \geq 3 \Rightarrow$ In both cases, an introspective equilibrium exists. Lastly, $\rho_{HA}^2 > \rho_{HB}^2$ and thus $\rho \in (\rho_{HB}^2, \rho_{HA}^2)$. This is the mirror image to above. HA chooses action H at level-2 while HB chooses action L at $k \geq 2$, ∴ an introspective equilibrium exists in all possibilities

Lastly, (N.4) $\rho_{HA}^1 > \rho_{HB}^1 > \rho_{LB}^1 > \rho_{LA}^1 > \rho_{HC}^1 > \rho_{LC}^1$ $(N.4) \rightarrow \sigma^0 = (H,H,L,H,L,L)$

(1) $\rho < \rho_{LC}^1 \Rightarrow \sigma^1 = (H,H,H,H,H,H)$
(2) $\rho \in (\rho_{LC}^1, \rho_{HC}^1) \Rightarrow \sigma^1 = (H,H,H,H,H,L)$
(3) $\rho \in (\rho_{HC}^1, \rho_{LA}^1) \Rightarrow \sigma^1 = (H,H,H,H,L,L)$
(4) $\rho \in (\rho_{LA}^1, \rho_{LB}^1) \Rightarrow \sigma^1 = (H,H,H,L,L,L)$
(5) $\rho \in (\rho_{LB}^1, \rho_{HB}^1) \Rightarrow \sigma^1 = (H,H,L,L,L,L)$
(6) $\rho \in (\rho_{HB}^1, \rho_{HA}^1) \Rightarrow \sigma^1 = (H,L,L,L,L,L)$
(7) $\rho > \rho_{HA}^1 \Rightarrow \sigma^1 = (L,L,L,L,L,L)$

Case (1) and (7) are already introspective equilibria for the same reason as in the demonstration in chapter 4. Case (2) is the same as in the other (N.), an introspective equilibrium exists. For case (3) the ones with stereotypes choose action L, the ones without stereotypes choose action H. At level-2 the everyone needs to be considered as $\mathbb{E}^2[m \mid I_j, G_j] = 1 - \alpha\gamma \coloneqq \rho_{IG}^2$ for all I, G. For $\rho > \rho_{IG}^2$ everyone chooses L in introspective equilibrium, for $\rho < \rho_{IG}^2$ everyone chooses H in introspective equilibrium. Case (4) for (N.4) works the same as at (N.3). At level-2 the expected values of HA and HC are equal as well as for LA and LC. It reduces to 4 different $\rho_{IG}^2 \rightarrow$ it is the same situation as in in the original paper $\Rightarrow$ an introspective equilibrium exists. Case (6) is again identical to the other rankings, the proof of this case is therefore omitted.

With this I've proven that under negative stereotypes, an introspective equilibrium always exists.

q.e.d.