Multidimensional Social Identities and Choice Behavior: The Pitfalls and Opportunities

Caroline Liqui-Lung [*]

Version: November 4, 2024
Click for Most Recent Version

When beliefs about success in a task are noisy, agents can improve decision making on average through biasing beliefs towards their welfare-maximizing task. I propose a model in which social identity serves as an instrument to mechanically bias beliefs, introducing an upward bias when a group is overrepresented among those successful, and a downward bias otherwise. The key insight of the model is that social context, i.e. data on group composition of successful people in each task, induces differences in the options to bias beliefs available to agents with different social identities. This induces differences in the propensity to choose a task across agents, which can make the representation of groups in the task a self-fulfilling prophecy. Moreover, exogenous changes in social context create externalities in choice behavior across traits, e.g. gender and ethnicity. I discuss the pitfalls and opportunities these externalities create for policy.

Keywords: Bounded Rationality, Social Identity, Choice Behavior, Diversity
JEL: D-81, D-91, Z-13

[*] University of Cambridge (cll64@cam.ac.uk). I would like to thank Olivier Compte, Philippe Jehiel, Roland Bénabou, Alessandra Casella, Debraj Ray, Jeanne Hagenbach, Glenn Ellison, Matthew Elliott and the participants of the Barcelona Summer Forum 2024, EEA Annual Congress 2023, THRED 2024, the Warwick Icebreaking Workshop 2024 and seminars at the University of Cambridge, UPF Barcelona, University of Bristol, the Paris School of Economics, the University of Leicester and the University of Liverpool for their useful comments and suggestions.

1 Introduction

Persistent inequalities along the lines of e.g. gender, ethnicity or social class are associated with lower social mobility and productivity, and a decrease in cohesion and trust in government and institutions. Such inequalities are fueled through the fact that individuals belonging to different social groups form different beliefs regarding own abilities, which results in different occupational and educational choices (Guyon and Huillery, 2021).[1] Although the underrepresentation of women and ethnic minorities in certain domains are predominantly considered as separate phenomena, data suggests this approach does not capture the entire picture. The 'Leaders and Daughters Global Survey 2017' documents how women's ambitions fall as they strive towards top leadership positions, where this downward trend is disproportionately strong for women belonging to ethnic minorities. Moreover, reactions to changes in social context provide additional puzzling evidence. For example, introducing a female-only math contest leads to an increase in the overrepresentation of Asian participants.[2] Moreover, these reactions are driven by ability. Artificially making gender salient in math tests predominantly affects the performance of girls who are strong in math, and not so much those who were not performing well to begin with (Steele, 2010).

The idea that self-evaluations rely on social comparison has been well-established.[3] A simple approach to explain the data could therefore be one where people mechanically bias their beliefs upwards when people like them are relatively overrepresented among those successful in a task, while they bias beliefs downwards when people like them are underrepresented. Yet, such a story implies a bias away from correct beliefs that can foster suboptimal decisions. Moreover, it cannot explain the reactions to changes in social context we observe in the data, and still leaves other questions unanswered. For example, how do people choose the groups they identify with?

I propose a model in which social identity is an instrument to mechanically bias beliefs, but agents have the option to repress this bias. In particular, it is well known that people find it difficult to objectively evaluate their own abilities (Mobius et al.,

[1] See literature on self-stereotyping, e.g. Bordalo et al. (2019), Coffman (2014), Flory et al. (2015) and Lippmann and Senik (2018).

[2] See data of the AT foundation and their Math Prize for Girls.

[3] See Section 2 for an extensive review of the literature on this topic.

2014). Moreover, such evaluations are affected by emotions, situational factors or recent feedback in a way that is beyond people's control.[4] When agents have such noisy beliefs, they can improve decision making on average when beliefs are distorted in the direction of their welfare-maximizing task. Meaning, individuals talented in a task would want to bias beliefs upwards to avoid being too pessimistic and not undertaking the task, while not talented individuals would want to bias beliefs downwards.

The key insight of the model is that social context, i.e. data on task allocation and group composition of successful people in a task, determines how an exogenously specified social type translates into a set of options to bias beliefs that differs across agents. Underlying ability then determines the bias agents adopt. For example, a male Asian student deciding whether to enter a Math Olympiad may identify with other male students, other Asian students or students who are both male and Asian. Data on those successful in the previous cohort shows male and Asian students were relatively overrepresented among those successful. Hence, all these identification strategies would translate into an upward bias of beliefs. Talented male Asian students will want to use this option to bias, but an upward bias would hurt male Asian students who are not talented. They would therefore wish to repress the option to bias. Female Asian students, on the other hand, would always want to use the option to bias beliefs. They will identify with Asian students to bias beliefs upwards when they are talented, and with female students to bias belief downwards when they are not talented.

Differences in the options to bias induce both a difference in the propensity to choose the task across social types, and a difference in mean competence. In the Math Olympiad example, female Western students will have the lowest propensity to enter the Olympiad, because they cannot bias beliefs upwards. Yet, they tend to be more successful on average. Male Asian students will have the highest propensity to enter the Olympiad, because they cannot bias beliefs downwards, but have the lowest on average success rate. I show how these effects can fuel differences in representation along the lines of gender and ethnicity. Under certain conditions, once particular traits start to prevail among those successful, an asymmetric outcome where agents with certain social types enter the task more often and hence prevail, becomes the only stable

[4] See e.g. Fiedler and Bless (2000), Ross and Nisbett (1991) and Elster (1996)

equilibrium outcome. Hence, the underrepresentation of certain traits in social context can become a self-fulfilling prophecy, even when agents are behaving optimally from an individual perspective. This can perpetuate differences in representation induced by historical factors that are no longer relevant. The first message of this paper is therefore that, if we want to address differences in choice behavior across social groups, taking care of discrimination and initial skill differences is not enough.

Moreover, the externalities in behavior along the lines of different traits that arise as a result of changes in social context in the model can explain the earlier mentioned patterns in the data. In a simple model with two traits, restricting the set of participants in a math contest to female students only takes away the option of not talented female Asian students to use the underrepresentation of female students to bias their beliefs downwards. As a result, they will now repress the bias all together, which will increase the participation of Asian students in the contest. Moreover, we can introduce stigma in the model by making the options to repress the bias along the stigmatized trait costly. This creates yet different externalities. When women are stigmatized in math contests, this makes the option for talented female Asian students to bias beliefs upwards costly. This decreases the participation of both female and Asian students in math contests, increasing differences in representation along the lines of gender, while decreasing them along the lines of ethnicity. Furthermore, when confidence affects performance, this explains why talented women particularly suffer from stereotype threat. Similarly, African American women may only be able to identify with their full type, as their experience in society is very different from white women or African American men (Crenshaw, 1991). When African American women are stigmatized and poorly represented in a given task, this will undermine the prevalence of women in general in that task, and, hence, negatively affect the option to bias beliefs upwards for all talented women. On the other hand, the ability to manipulate an identity, such as the concept of 'racial passing' (Qian and Nix, 2015), or to self-identify, like adopting a particular fashion style, increases the options for agents to bias beliefs, creating opposite externalities. In general, these externalities create both pitfalls and opportunities for policy aiming to achieve social diversity, and the second message of the paper is that a multidimensional view on social identity is crucial for effective policy design.

The first contribution of the paper is that I shed light on the instrumental role of both the use of social identity in decision making, and the choice of group agents identify with. Contrary to what has been done in the literature (e.g. Akerlof and Kranton (2000), Shayo (2009) and Carvalho and Pradelski (2022)), I do this without assuming social identity directly affects preferences or performance, nor that it is relevant in a Bayesian sense. Instead, I show how social identity can be an instrument to optimally manage noisy confidence about chances of success in a task. Furthermore, identity-contingent behavior is not guided by exogenously imposed stereotypes or norms, but arises endogenously in equilibrium through social context.

The second contribution of the model is that it provides a possible explanation for why one-dimensional policy measures, such as those centered around gender alone, have unwanted spillover effects on the representation of other traits (Cassan and Vandewalle, 2021). Furthermore, the analysis of the externalities in behavior across different traits provides novel insights into why one-dimensional affirmative action policy cannot achieve equal representation in multiple dimensions (Carvalho et al., 2023) and novel ideas as to how to design more effective multidimensional policy approaches.

The third contribution is that the model provides an intuition for why persistent differences in choice behavior arise particularly along the lines of traits like gender, ethnicity and social class. Such traits are often stigmatized in society and costly to manipulate. This makes it easier for an asymmetric equilibrium to exist, and reinforces the asymmetries in choice behavior that can persist. On the other hand, an asymmetric equilibrium outcome is difficult to obtain along the lines of traits that are easy to adopt, such as a particular fashion style. To my knowledge, it is the first equilibrium model that has something to say regarding this type of equilibrium selection.

Finally, the paper makes several contributions in a methodological sense. Modeling bounded rationality through endowing agents with a limited family of rules is in the spirit of Compte and Postlewaite (2019). The novelty in this paper is that the family of rules represents a family of belief-formation strategies that is determined endogenously as a result of social context. Agents adopt the rule that maximizes their ex-ante expected welfare. Like in Compte and Postlewaite (2004) and Brunnermeier and Parker (2005), agents adopt systematically biased beliefs when this enhances expected utility.

A different way to understand the approach is through the lens of the literature on motivated beliefs and present bias (e.g. Benabou and Tirole (2011), Benabou and Tirole (2006), O'Donoghue and Rabin (1999)). There, agents are confronted with biased decision rules and commit to a strategy (e.g. distorting memory) that increases ex-ante welfare. In this model, agents are not confronted with a systematically biased decision rule, but one that incorporates errors. They commit to a strategy, i.e. adopting a (possibly misspecified) model of the world about to influence of certain traits on success, that maximizes ex-ante expected utility. This also relates the paper to the experimental literature documenting how agents correct for one bias, i.e. present bias or self-image, with another, i.e. self-deception or information avoidance.[5]

Furthermore, I use a static solution concept analyzing the fixed points in social context induced by the individually optimal strategies. In the literature on subjective and/or misspecified beliefs, such as Esponda and Pouzo (2016), Fudenberg and Levine (1993) or Spiegler (2016), equilibrium beliefs are consistent with observational feedback, ensuring they are closest to the truth. In this model, equilibrium beliefs are disciplined through a fitness criterion, allowing agents to make decisions that are better aligned with welfare maximization. The model raises therefore the question whether there is legitimacy for using Bayesian updating as a belief-formation rule when perceptions are noisy to begin with and agents do not have the tools to correct for this noisiness. Finally, each identification strategy could be interpreted as a model of the world describing what traits agents believe to be relevant for ability. This relates the paper to the literature on narratives and model selection (e.g. Schwartzstein and Sunderarm (2021) and Eliaz and Spiegler (2020)). Contrary to this literature, in this paper, model selection is not driven by a likelihood criterion. Instead, agents adopt the model of the world that 'works best of them' in terms of maximizing ex-ante welfare.

The paper is organized as follows. Section 2 reviews the literature motivating social behavior in this model. Section 3 presents the model. Section 4 analyzes individual and aggregate choice behavior. Section 5 discusses what fosters persistent identity-driven choices. Section 6 discusses externalities and policy implications. Section 7 concludes, and Appendix 1 presents the formal proofs.

[5] See e.g. Exley and Kessler (2019), Van de Weele et al. (2022) and Saccardo and Serra-Garcia (2023)

2 Social Influences on Choice Behavior

This paper builds on the idea that people pay attention to behavior and outcomes of others like them when evaluating their optimal choice of task. Moreover, the type of social cues people use in decision making are determined by social context and people's underlying ability. These ideas originated in social psychology and evolutionary biology. With social comparison theory, Festinger (1954) first popularized the idea that individuals have a primitive drive to compare themselves to others when evaluating their own opinions and abilities. Hogg and Grieve (1999) argue people identify with groups to enhance self-confidence and to reduce subjective uncertainty. In the process of depersonalization, which is associated with social identification, individual and concomitant unshared beliefs, attitudes, feelings, and behaviors are replaced by an in-group prototype that prescribes shared beliefs, attitudes, feelings and behaviors. Seligman (2006) shows how people can interpret numerous failures from others like them as evidence they will fail as well. Steele (2010) discusses how the psyche of the individual gets damaged by repeated exposure to bad images of their group projected in society, leading to low self esteem, low motivation, and self doubt.

Henrich (2016) discusses how natural selection has shaped our brains to acquire information from the behavior of others. These learning instincts are efficient, where people focus on others with success, prestige and age, i.e. having more experience. Moreover, people learn from experience which self-similarity cues, such as gender and ethnicity, allow them to more effectively acquire the skills, norms and preferences that make them successful in their particular environment. Through experience, they learn when such cultural learning should overrule their own direct experiences, and vice versa. Because experience is driven by a persons' underlying abilities and characteristics, different people can end up paying attention to different cues in similar choice settings. This is in line with Pronin et al. (2004), showing how women who are strong in math actively disidentify with female aspects associated with a negative gender-math stereotype, while Steele et al. (2002) shows less talented female undergraduates report they believe they have weak abilities in math and science because of their gender.

These theories are furthermore supported by neurological evidence. Murden (2020) reviews the evidence for a mirror neuron system that continuously mirrors our behavior, intentions and beliefs onto others' brains and vice versa. Using fMRI technology, Reynolds Losin et al. (2012) shows how, neurologically speaking, people find it more rewarding to mimic others with their own gender. Such learning instincts are acquired at a very young age, where biases in learning from same-sex versus opposite-sex models emerge even before children develop a gender identity (Henrich, 2016). The resulting behavior is largely determined by the automatic part of our brain, outside of our awareness (Banaji and Greenwald (2016) and Murden (2020)) and it can be difficult to repress our automatic imitative instincts (Ross and Nisbett (1991) and Henrich (2016)). Finally, Ross and Nisbett (1991) shows how, especially in settings characterized by ambiguity and uncertainty, such as how hard a task is, or how capable one is, social influence plays an important role.

Social influences on self-perception affect choice behavior. For example, in Smith et al. (2007), people completing a high stereotype-threat test report decreased task interest. Davies et al. (2002) shows how the combination of decreased enjoyment and diminished self-confidence explains why women experiencing stereotype threat report less interest in math and science fields and weaker leadership aspirations compared to men or non-threatened women. Similarly, in Banaji and Greenwald (2016), implicit associations picked up from social context affect our behavior, such as the intellectual pursuits we select, and Perry et al. (2003) discusses how people tend to protect themselves from stereotype threat by ceasing to care about the domain in which the stereotype applies. Finally, Oh (2023) shows how Indian workers are willing to forego substantial payments to avoid tasks that are associated with other castes.

3 The Model

3.1 The Environment

Agents - I consider a society with $i = 1, ..., N$ agents, with N arbitrarily large. Agents are first of all described by an *ability type*, that is captured by the continuous variable $\alpha_i \in [0, 1]$. This *ability type* is fixed for each individual and is distributed over the population following a distribution $f_\alpha \in [0, 1]$. Secondly, each agent has a multidimensional *social type* that represents for example their gender, ethnicity or social class, and is public information. To simplify the exposition of the model, let $\Theta_i = (\theta_i^k)_{k \in \{A,B\}}$ be the social type of agent i, where each θ_i^k is a binary *trait* with realizations $\theta^k \in \{0, 1\}$. Let $\Theta \in \{11, 10, 01, 00\}$ be a possible realization of the social type Θ_i. I let p_{θ^k} be the fraction of the population with trait $\theta_i^k = \theta^k$, while p_Θ is the fraction of the population with social type $\Theta_i = \Theta$. To isolate the mechanism through which social context affects choice behavior in this model, I assume ability types α_i and social types Θ_i are independently distributed over the population.[6]

Action Space - Each agent chooses an action $a_i \in \{C, NC\}$, where C and NC represent classes of tasks of respectively a *Competence-Driven* and a *Non-Competence-Driven* type. The outcome of action a_i can be either *'success'* or *'failure'*, which is represented by the binary variable $Y_i \in \{0, 1\}$. Agents derive utility from a successful outcome, independent of their choice of task. Hence, their utility function can be denoted by $U_i = Y_i$. The probability of success for a *Competence-Driven* task depends on an agent's *ability type*, such that for each agent i, the probability of a successful outcome $Y_i = 1$ conditional on choosing this task is given by,

$$p(Y_i = 1 | a_i = C) = \alpha_i$$

For simplicity, I assume the *Non-Competence-Driven* task has a probability of success $\gamma \in [0, 1]$ that is known and the same for all agents. Therefore, for all i,

$$p(Y_i = 1 | a_i = NC) = \gamma$$

More generally, γ can be interpreted as the attractiveness of the *Non-Competence-Driven* task relative to the *Competence-Driven* task. To transmit the main insights of the model in the simplest way, I assume α_i and γ are fixed for each agent.[7]

Social Context - Agents have access to public data about behavior of others in society. First, I introduce *social identity cues* $\pi_{\theta^k} \in [0, 1]$, with $\theta^k \in \{0, 1\}$ and $k \in \{A, B\}$. These cues are statistics agents observe about the performance or prevalence of others

[6] The model can account for multiple and for non-binary observable traits. See Section 3.4 for a discussion on the implications of correlated observable traits. See Liqui Lung (2022) for a discussion on what happens when α_i and Θ_i are correlated.

[7] See Liqui Lung (2022) for a discussion on the case in which these variables vary over time.

with trait θ^k among those undertaking or successful in the *Competence-Driven* task. The '*social context*' of the population is defined as the vector $\Pi = (\pi_{\theta^k})_{k\in\{A,B\},\theta^k\in\{0,1\}}$. Secondly, I define $\overline{\pi}_{\theta^k} \in [0,1]$ as the *benchmark* that would arise when individuals with different *social types* have the same choice behavior and success rates given the distribution of *social types* in the population. To illustrate these cues, consider the following example.

Example - In many real-life settings, people only have access to data about a pool of successful individuals. Moreover, research shows people are more inclined to learn from and copy successful individuals (Henrich, 2016). Let $\mathcal{N}_{C,\theta^k} = \{i \in N, \theta_i^k = \theta^k, a_i = C\}$ be the set of agents with $\theta_i^k = \theta^k$ that have chosen the *Competence-Driven* task. Let $\mathcal{N}_C = \{i \in N, a_i = C\}$ be the set of all agents that have chosen the *Competence-Driven* task, which implies $\mathcal{N}_{C,\theta^k} \subset \mathcal{N}_C$. Agents could calculate the following statistics,

$$\pi_{\theta^k} = \frac{\sum_{i\in\mathcal{N}_{C,\theta^k}} Y_i}{\sum_{i\in\mathcal{N}_C} Y_i}$$

for all $k \in \{A,B\}$ and $\theta^k \in \{0,1\}$. These are the fractions of successful individuals with trait $\theta_i^k = \theta^k$ among all successful individuals that have chosen the *Competence-Driven* task. The appropriate benchmarks are equal to $\overline{\pi}_{\theta^k} = p_{\theta^k}$.

Noisy Perceptions - Evaluating α_i is difficult, and factors, such as recent feedback or emotions, can make agents momentarily too optimistic or pessimistic. I introduce such momentary noise in decision making by assuming that, when choosing action $a_i \in \{C, NC\}$, agents do not have access to α_i, but only to a noisy perception $\hat{\alpha}_i$. To show a systematic bias is not the mechanism that drives the results in this model, I assume this noisy perception is unbiased, such that it sometimes tilts the decision in favor of the *Competence-Driven* task and sometimes against it.

ASSUMPTION 1: *When choosing action $a_i \in \{C, NC\}$, agents only have access to a noisy perception $\hat{\alpha}_i$ of α_i stemming from a distribution $g_{\alpha_i,\hat{\alpha}} \in [0,1]$ with $E(\hat{\alpha}_i) = \alpha_i$.*

Strategies - The second key assumption in the model is that agents do not have the tools to fully correct for the noise in their perception $\hat{\alpha}_i$.

ASSUMPTION 2: *Agents cannot fully exploit the structure of the model, and only have limited abilities to correct for the noise in their perception $\hat{\alpha}_i$.*

This aspect of bounded rationality is modeled through direct restrictions on the strategy set. In particular, I model agents that have a natural tendency to look at others when evaluating their optimal choice of task. Agents can either *Repress* this urge to look at others, or choose on which particular subgroup of agents they want to focus. The multi-dimensionality of the *social type* implies agents can define "others like them" in a flexible way, where they can focus on others that have either one of their traits or their entire social type in common. To simplify the exposition of the model, I assume agents can focus on others with either their trait θ^A, or their trait θ^B. Hence, agents choose a strategy $\sigma_i \in \{R, \theta^A, \theta^B\}$. In Section 4.3, I introduce the option for agents to look at others with the same social type Θ_i.

Response Function - Because α_i and Θ_i are independently distributed over the population, '*social context*' is not relevant to agents in a Bayesian sense. Instead, I introduce the option to agents to use their *social type* Θ_i to mechanically bias decision making in a direction contingent on the trait they focus on. To capture the direction and strength of this bias in a particular social context Π, I introduce a *response function* $\eta_{\pi_{\theta^k}, \overline{\pi}_{\theta^k}} : (\pi_{\theta^k}, \overline{\pi}_{\theta^k}) \rightarrow \mathcal{R}$. I will then investigate how properties of this *response function* can be conductive to the phenomenon I mean to describe. For any value $\pi_{\theta^k}, \overline{\pi}_{\theta^k} \in [0, 1]$, η is non-decreasing, such that

$$\eta(\pi_{\theta^k}, \overline{\pi}_{\theta^k}) = \begin{cases} > 1 & \text{if } \pi_{\theta^k} > \overline{\pi}_{\theta^k} \\ 1 & \text{if } \pi_{\theta^k} = \overline{\pi}_{\theta^k} \\ < 1 & \text{if } \pi_{\theta^k} < \overline{\pi}_{\theta^k} \end{cases}$$

This implies the response function is larger than one when the *social identity cue* exceeds the benchmark, while it is smaller than one when the cue is smaller than the benchmark. The strength of the bias is determined by how much the social identity cue π_{θ^k} deviates from the benchmark $\overline{\pi}_{\theta^k}$. To simplify notation, I let

$$\eta_{\sigma_i} = \begin{cases} \eta(\pi_{\theta_i^k}, \overline{\pi}_{\theta_i^k}) & \text{when } \sigma_i \in \{\theta^A, \theta^B\} \\ 1 & \text{when } \sigma_i = R \end{cases}$$

A Model of Belief Formation - One interpretation of the model is that the instrument $\sigma_i \in \{R, \theta^A, \theta^B\}$ mechanically alters the agent's subjective belief about her chances of success.[8] I let the strategies $\sigma_i \in \{R, \theta^A, \theta^B\}$ give rise to a limited family of belief formation processes, where each strategy translates into a possible subjective belief $\hat{p}_i^{\sigma_i} \in [0,1]$ about α_i, such that,

$$\hat{p}_i^{\sigma_i} = \begin{cases} \hat{\alpha}_i & \text{if } \sigma_i = R \\ \eta_{\sigma_i}\hat{\alpha}_i & \text{if } \sigma_i \in \{\theta^A, \theta^B\} \end{cases}$$

The agent's subjective belief can take three values; $\hat{p}_i^R$, $\hat{p}_i^{\theta^A}$ and $\hat{p}_i^{\theta^B}$. With a subjective Bayesian interpretation in mind, the strategy $\sigma_i = R$ represents a model of the world in which *social types* and *ability types* are uncorrelated, and agents naively follow their noisy perception $\hat{\alpha}_i$. The strategies $\sigma_i \in \{\theta^A, \theta^B\}$ represent a model of the world in which *social types* and *ability types* are correlated. This biases agents' noisy perception $\hat{\alpha}_i$ with their *social identity cue* π_{θ^k} in a direction contingent on their choice of trait θ_i^k. When this trait is more successful in the current context, this belief-formation rule leads to an optimistic interpretation of $\hat{\alpha}_i$, while this leads to a pessimistic interpretation when this trait is socially less successful. When $\pi_{\theta^k} = \overline{\pi}_{\theta^k}$ for $k \in \{A, B\}$, the beliefs $\hat{p}_i^{\sigma_i}$ are equivalent for $\sigma_i \in \{R, \theta^A, \theta^B\}$.

A Model of Choice - A second interpretation of the model is that agents have the option to use social context Π to alter choice in a direction contingent on their social type Θ_i. Formally, subjective expected utility maximization implies the agent is effectively comparing thresholds three $\gamma_i^{\sigma_i} \in [0,1]$, such that agent i chooses $a_i = C$

[8] The cue π_{θ^k} could affect belief formation through both a bias in the prior and posterior. To maintain flexibility regarding the channel through which such a bias is obtained, I do not propose a particular functional form, nor root belief formation in a specific subjective Bayesian model.

if and only if $\hat{\alpha}_i > \gamma_i^{\sigma_i}$, where

$$\gamma_i^{\sigma_i} = \begin{cases} \gamma & \text{when } \sigma_i = R \\ \frac{\gamma}{\eta_{\sigma_i}} & \text{when } \sigma_i \in \{\theta^A, \theta^B\} \end{cases}$$

The strategies $\sigma_i \in \{\theta^A, \theta^B\}$ imply therefore that agents inflate or deflate the threshold for $\hat{\alpha}_i$ above which they think they are 'good enough' to undertake the *Competence-Driven* task. The choice set $\gamma_i^{\sigma_i} \in \{\gamma, \frac{\gamma}{\eta_{\theta^A}}, \frac{\gamma}{\eta_{\theta^B}}\}$ can be different for agents with different social types Θ_i, which will be the key driver of the equilibrium results.

Final Note - The limited strategy space $\sigma_i \in \{R, \theta^A, \theta^B\}$ implies agents cannot compare all functions of $\hat{\alpha}_i$ and π_{θ_i}. This aspect of bounded rationality should be considered a modeling device that helps to keep the model parsimonious. The key objective here is to show the difference with a Bayesian model, by analyzing whether, when agents do not have the tools to correct for all types of noise, this may open the door for them to use cues that are irrelevant in a Bayesian sense, but could still improve decision making. We can therefore consider larger strategy spaces, as long as Assumption 2 remains preserved.

3.2 The Solution Concept

Choices of strategies σ_i affect choices of tasks a_i. This leads to outcomes Y_i that induce cues π_{θ^k} that in turn affect choices of strategies σ_i. To tractably capture the fixed points in this dynamic process, I use a static solution concept.

Individual Optimality - The adoption of a strategy $\sigma_i \in \{R, \theta^A, \theta^B\}$ affects behavior through the ex-ante probability with which an agent chooses the *Competence-Driven* task over all possible realizations of the noisy perception $\hat{\alpha}_i$. Let $\Phi(\alpha_i, \Theta_i, \sigma_i, \Pi) \in [0, 1]$ denote this ex-ante induced probability for an agent of type $\{\alpha_i, \Theta_i\}$ playing strategy σ_i given a social context Π. For a given distribution $g_{\alpha_i, \hat{\alpha}}$, this probability is equal to,

$$\Phi(\alpha_i, \Theta_i, \sigma_i, \Pi) = P(\hat{p}_i^{\sigma_i} > \gamma | \alpha_i, \Theta_i, \Pi) \equiv P(\hat{\alpha}_i > \gamma_i^{\sigma_i} | \alpha_i, \Theta_i, \Pi)$$

The expected pay-off for agent i of type $\{\alpha_i, \Theta_i\}$ when choosing strategy $\sigma_i \in \{R, \theta^A, \theta^B\}$

given Π over all possible realizations of $\hat{\alpha}_i$ is,

$$V_i(\sigma_i|\alpha_i, \Theta_i, \Pi) = \alpha_i \Phi(\alpha_i, \Theta_i, \sigma_i, \Pi) + \gamma(1 - \Phi(\alpha_i, \Theta_i, \sigma_i, \Pi))$$

I then define individual optimality as follows.

DEFINITION 1 (Individual Optimality): *Strategy σ_i^* is optimal for agent i given Π if,*

$$\sigma_i^* = \underset{\sigma_i \in \{R, \theta^A, \theta^B\}}{\text{argmax}} \; V_i(\sigma_i|\alpha_i, \Theta_i, \Pi)$$

The optimal strategy σ_i^* maximizes therefore an agent's expected utility on average over all possible realizations of $\hat{\alpha}_i$ given their type $\{\alpha_i, \Theta_i\}$ and social context Π. I assume agents compare $V_i(\sigma_i|\alpha_i, \theta_i, \Pi)$ and choose their strategy according to Definition 1. We can interpret this assumption in the spirit of motivated beliefs (Benabou and Tirole, 2011) or a two-selves approach (O'Donoghue and Rabin, 1999). In a first stage, the (sophisticated) agent knows the true ability type α_i and commits to a strategy σ_i, while in a second stage, the (impulsive) agent only has access to $\hat{\alpha}_i$ and chooses their action a_i given the rule implied by the strategy chosen by the (sophisticated) agent in the first stage. Alternatively, the behavior can be motivated in line with the literature presented in Section 2, where agents learn their optimal strategy from their own experience with similar choices of tasks through for example reinforcement learning or a sampling process[9]. The true probability of success α_i determines the feedback agents observe in such a process, which enables them to learn whether it is optimal to *Repress* or focus on a certain subgroup without precise knowledge of the relationship between the choice of strategy σ_i, choice of task a_i and the observed outcome Y_i. The fitness of a strategy is determined by an agent's type and social context, and, because the set of strategies is small, it is easy for agents to compare their strategies.[10]

[9] The dynamic story underlying the reduced-form analysis is that agents make related *Competence-Driven* choices throughout their lifetime. For example, early in life they choose whether to 'undertake a math-related major', while later in life they choose whether to 'pursue a STEM career'.

[10] It may seem plausible that, if agents are able to learn their optimal strategy σ_i conditional on α_i, they should also be able to retrieve their true value of α_i from this optimal strategy. This line of thought is nevertheless driven by the simplification of the model in which α_i and γ_i are fixed over

Population Equilibrium - Let $\sigma \in \{R, \theta^A, \theta^B\}^N$ be a profile of strategies σ_i. Because N is arbitrarily large, for a given $(\eta, f_\alpha, g_{\alpha_i,\hat{\alpha}}, \gamma, (p_{\theta^k})_{k\in\{A,B\}})$, each profile of strategies σ and social context Π generate choices a_i and outcomes Y_i that in turn generate an induced social context $\widetilde{\Pi}(\sigma, \Pi) = (\tilde{\pi}_{\theta^k}(\sigma, \Pi))_{k\in\{A,B\},\theta^k\in\{0,1\}}$, where each $\tilde{\pi}_{\theta^k}(\sigma, \Pi) \in [0, 1]$.

Example - When agents process the fraction of successful agents with type θ^k among all those successful in the *Competence-Driven* task, the induced social identity cue $\tilde{\pi}_{\theta^k}(\sigma, \Pi)$ is equal to,

$$\tilde{\pi}_{\theta^k}(\sigma, \Pi) = \frac{p_{\theta^k} \int \alpha \Phi(\alpha, \Theta, \sigma, \Pi) f(\alpha) d\alpha}{\sum_\Theta p_t \int \alpha \Phi(\alpha, \Theta, \sigma, \Pi) f(\alpha) d\alpha}$$

An population equilibrium in the model can now be defined as follows.

DEFINITION 2 (Population Equilibrium): *For each* $(\eta, f, g_{\alpha_i,\hat{\alpha}}, \gamma, (p_{\theta^k})_{k\in\{A,B\}})$*, a pair* $\{\sigma, \Pi\}$ *is a population equilibrium if* σ *is optimal given* Π*, and when,*

$$\Pi = \widetilde{\Pi}(\sigma, \Pi)$$

In other words, a population equilibrium is a fixed point in *social context* Π when all agents play their individually optimal strategy σ_i^* given Π. This solution concept is in line with the view that the optimal strategy σ_i^* arises from a learning process that operates faster than the dynamics in social context Π, where the learning of the optimal strategy happens during the lifetime of an agent through her experience with similar tasks, while changes in social context arise from agents belonging to different generations making a specific choice of task once in their lifetime.

4 Choice Behavior

4.1 The Individual Level

Using Definition 1, we can provide insights regarding how agents determine which trait of their social type they focus on in decision making as a function of their ability type

the lifetime of an agent. In Liqui Lung (2022), I discuss how the results are robust when α_i and/or γ vary and when learning would be imperfect.

and social context. I illustrate these insights with the following example.

Example - Consider a cohort of students that choose whether to undertake a large-scale math competition (C) or a generic task (NC). They observe a list of students from last year's competition that managed to qualify for the international final, and can see whether students were male or female and whether they have a Western or Asian last name. Assume female students and students with a Western last name were underrepresented among those that qualified. Let $Gender \in \{M, F\}$ and let the origin of a last name be denoted by $Name \in \{W, A\}$. Hence, agents can calculate the social identity cues π_{θ^k} capturing how many students with a *Gender* or *Name* are among all those that were successful last year. Furthermore, for simplicity, assume male students and students with an Asian last name were overrepresented to the same degree, such that $\eta_M = \eta_A$. To further simplify the discussion, I will refer to students with an ability type $\alpha_i > \gamma$ as *talented*, while I will refer to those with an ability type $\alpha_i < \gamma$ as *not talented.*

A first important insight is that momentary noise in the perception $\hat{\alpha}_i$ has an asymmetric effect on decision making, and each type of student is prone to making a different type of error. Students that are *not talented* can be too optimistic on the day of decision making, and decide to enter even though this is not optimal. This is what I will refer to as a Type-I error. Yet, when these students are too pessimistic, it will not affect their decision making. This is vice versa for *talented* students. When these students are too optimistic, it does not affect their decision making, while, when they are too pessimistic, they may not enter despite the fact that this would have been their welfare-maximizing choice. This is what I will refer to as Type-II error.

Figure 1 shows the thresholds $\gamma_i^{\sigma_i}$ that follow from the strategies $\sigma_i \in \{Gender, Name, R\}$ for a male student with a Western last name. Because male students were overrepresented among those successful last year, focussing on this subgroup would deflate the threshold for $\hat{\alpha}_i$ above which he thinks he is 'good enough', while focussing on those with a Western last name would inflate this threshold. The arrows show the induced probabilities $\Phi(\alpha_i, \Theta_i, \sigma_i, \Pi)$ with which the agent chooses to enter the math competition for each strategy σ_i given the social context Π.

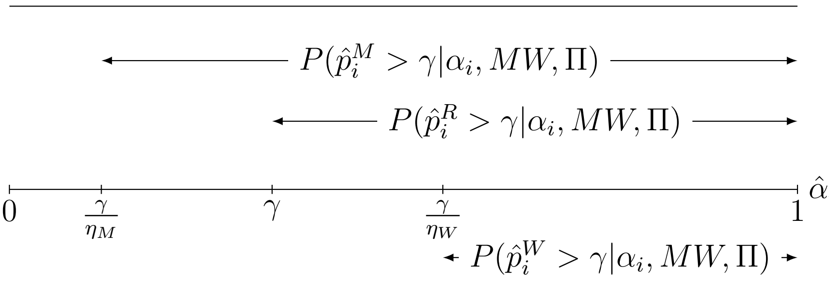
$P(\hat{p}_i^M > \gamma|\alpha_i, MW, \Pi)$
$P(\hat{p}_i^R > \gamma|\alpha_i, MW, \Pi)$
$\hat{\alpha}_i$
0
$\frac{\gamma}{\eta_M}$
γ
$\frac{\gamma}{\eta_W}$
1
$P(\hat{p}_i^W > \gamma|\alpha_i, MW, \Pi)$

Figure 1: The induced probabilities with which a Male student with a Western last name undertakes the math competition for different choices of σ_i

Consider first a *talented* male student with a Western last name. To maximize expected utility, he should enter the competition. Yet, he is prone to making a Type-II error when, at the moment of decision making, he is too pessimistic. When he focusses on the social identity cue regarding successful students with a Western last name, he will inflate the threshold for $\hat{\alpha}_i$ above which he enters the competition. This increases his chances of making a Type-II error. On the other hand, when he focusses on successful male students, he deflates the threshold for $\hat{\alpha}_i$ above which he enters the competition. This enables him to minimize the likelihood he makes a Type-II error over all possible realizations of his noisy belief $\hat{\alpha}_i$. This student can therefore improve decision making on average when he believes the relative overrepresentation of male students is a sign he will be more successful in the competition as well, while disregarding the fact that students with a Western last name are underrepresented among those currently successful. This is exactly vice versa when he is *not talented.* He can minimize the likelihood of making a Type-I error by learning to believe that having a Western last name decreases his chances of success in the competition, while disregarding the fact that male students are overrepresented among those that qualified.

These insights can be generalized as follows. Define $\overline{\eta}_\Theta = \max_k \eta_{\theta^k}$ and $\underline{\eta}_\Theta = \min_k \eta_{\theta^k}$. Similarly, let $\overline{\kappa}_\Theta = \text{argmax}_k\, \eta_{\theta^k}$ and $\underline{\kappa}_\Theta = \text{argmin}_k\, \eta_{\theta^k}$.

PROPOSITION 1 (Individually Optimal Belief Formation): *The individually optimal strategies σ_i^* given an agent's type $\{\alpha_i, \Theta_i\}$ and a social context Π are the following:*

	$\alpha_i > \gamma$	$\alpha_i < \gamma$
$\overline{\eta}_\Theta > 1$	$\overline{\kappa}_\Theta$	R
$\underline{\eta}_\Theta < 1$	R	$\underline{\kappa}_\Theta$

Proposition 1 shows how agents endogenously determine which dimension of their social type affects choice behavior as a function of their exogenously specified ability type and social context. Agents can use social identity cues to bias their decision making towards a certain task. When they focus on the cue that best biases their decision making in the direction of their welfare-maximizing task, they can improve decision making on average over all possible realizations of their noisy perception $\hat{\alpha}_i$.

Talented individuals focus on the trait that best boosts their confidence. This is in line with literature on enhancing confidence being one of the motivations for social identification (see e.g. Hogg and Grieve (1999), Akerlof (2016a) or Akerlof (2016b)). Individuals that are *not talented* will instead focus on the trait that makes them most pessimistic. This is in line with the concept of a psychological immune system, that consists of a series of processes that adjusts beliefs to protect us from threats to one's sense of self (Rosenzweig, 2016). One of these processes is attributing negative outcomes to group aspects rather than individual aspects. Finally, Proposition 1 is in line with Pronin et al. (2004), showing how women who are strong in math actively disidentify with female aspects associated with a negative gender-math stereotype, while Steele et al. (2002) shows less talented female undergraduates report they believe they have weak abilities in math and science because of their gender.

4.2 The Aggregate Level

4.2.1 Potential to Improve Decision Making

Proposition 1 shows how agents can use social identity cues to decrease the likelihood of making mistakes in decision making due to the noise in their perception $\hat{\alpha}_i$. The multidimensionality of social types allows agents to have a more flexible interpretation of their social context. In this section, I show under which conditions this flexibility enhances the agent's potential to improve decision making.

Example - As we saw in the previous section, a *not talented* male student with a Western last name can decrease the likelihood of making a Type-I error by focussing on other students with a Western last name, while, when he is *talented*, he can decrease the likelihood of making a Type-II error by focussing on other male students. A male student with an Asian last name can nevertheless only bias his noisy perception $\hat{\alpha}_i$ upwards, no matter the trait he focusses on. Therefore, this student can only minimize the likelihood of making a Type-II error, while, when he is *not talented*, the best he can do is to *Repress* the urge to look at others. This is vice versa for a female student with a Western last name. Table 1 shows for each social type whether they are able to reduce the likelihood of making a Type-I respectively Type-II error.

Θ	Type I error	Type II error
MA	No	Yes
MW	Yes	Yes
FA	Yes	Yes
FW	Yes	No

Table 1: The potential to improve decision making for each realization of the social type Θ

Table 1 shows male students with an Asian last name and female students with a Western last name can only decrease the likelihood of making one type of error, while female students with an Asian last name and male students with a Western last name can decrease the likelihood of making both types of mistake. These students are therefore on average more likely to choose their welfare-maximizing action.

In general, we can divide the set of social types T into two different categories. These different categories play a key role in determining aggregate choice behavior.

DEFINITION 3: *Given a social context* Π, *social types* Θ *can be categorized as follows:*

- *A social type* Θ *is* ***mixed****, when* $\underline{\eta}_{\Theta} < 1 < \overline{\eta}_{\Theta}$
- *A social type* Θ *is* ***one-sided****, when* $\underline{\eta}_{\Theta} > 1$ *or* $\overline{\eta}_{\Theta} < 1$

In other words, an agent with a *mixed* social type belongs to the socially more successful group according to one trait, but to the socially less successful group according to the other trait. An agent with a *one-sided* social type belongs to either the socially more or socially less successful group according to both traits. Using this definition, we can show one of the main insights of the paper.

PROPOSITION 2 (Potential to Improve Decision Making): *Asymmetry* $\underline{\pi}_{\theta^k} \neq \overline{\pi}_{\theta^k}$ *along the lines of at least two observable traits* θ^A *and* θ^B *leads to inequalities in the potential to improve decision making across the different social types* Θ*. Specifically, agents with a mixed social type will have on average a higher expected pay-off* $V_i(\sigma_i^*|\alpha_i, \Theta_i, \Pi)$ *than agents with a one-sided social type.*

Where Proposition 1 shows how agents can limit the adverse effects of their noisy perception on decision making with the use of social identity cues, Proposition 2 shows how the multidimensionality of a social type only reinforces this potential to improve decision making for agents with *mixed* social types. This creates a disadvantage for agents with *one-sided* social types relative to agents with a *mixed* social type, and hence an inequality in expected utility. This result is different from the literature on intersectionality, where the effect of multidimensional social identities operates through the adding up of advantages or disadvantages of one's group being over- respectively underrepresentation (see e.g. Yuval-Davis (2006) and Crenshaw (1991)). In this model, it is not under- or overrepresentation per se that determines whether agents are advantaged or disadvantaged. Both situations provide a tool to decrease the likelihood of making a particular type of error. The multidimensionality of a social type in this model affects decision making through the flexibility agents have to use this tool to their advantage. The larger this flexibility, the better agents will be able to cope with the potentially negative effects of the noise they are subject to. Despite this possibly counterintuitive result, we will see that the dynamics this individual behavior induces at the aggregate level are in line with what we see in the data.

Note - Proposition 2 implies there is no inequality between agents with the socially more successful and the socially less successful *one-side* social type. This should be considered as the result of simplifications made in the model to isolate the particular mechanism through which the results are obtained in this paper. Such an inequality will arise if we either assume beliefs have a direct effect on performance, such as in Compte and Postlewaite (2004), or when we assume choosing the *Competence-Driven* task more often affects ability itself. For example, choosing STEM-related activities repeatedly can lead to a better development of the abilities relevant for these activities, and an acquired taste for STEM-related careers. This can eventually lead to higher expected earnings.

4.2.2 Externalities in Aggregate Choice Behavior

In the following, I illustrate how differences in the potential to improve decision making given a social context Π can induce differences in choice behavior and average success rates across a priori identical agents with different types (α, Θ_i).

Example - Figure 2 shows the induced probabilities with which students choose to enter the math competition conditional on their type and the earlier described social context. All *talented* male students and all *talented* students with an Asian last name focus on the social identity cue based on the representation of these respective traits among those previously qualified. Their probability of entering the competition is represented by the top arrows. *Talented* female students with a Western last name and *not talented* male students with an Asian last name learn to *Repress* the use of social identity cues in decision making. Their induced probabilities to enter the competition are represented by the middle two arrows. Finally, *not talented* female students and students with a Western last name focus on the social identity cue based on the representation of these respective traits among the successful students. Their probability of entering the competition is represented by the lower two arrows.

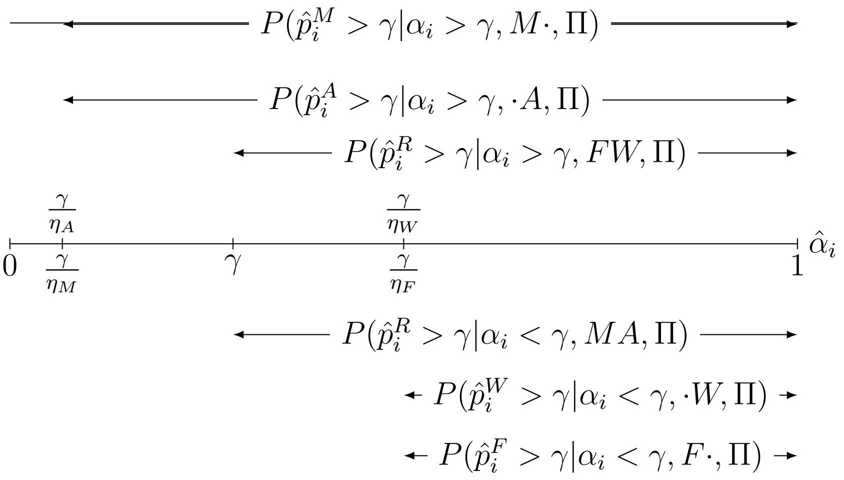
$P(\hat{p}_i^M > \gamma | \alpha_i > \gamma, M\cdot, \Pi)$
$P(\hat{p}_i^A > \gamma | \alpha_i > \gamma, \cdot A, \Pi)$
$P(\hat{p}_i^R > \gamma | \alpha_i > \gamma, FW, \Pi)$
$\frac{\gamma}{\eta_A}$
$\frac{\gamma}{\eta_W}$
0
$\frac{\gamma}{\eta_M}$
γ
$\frac{\gamma}{\eta_F}$
1
$\hat{\alpha}_i$
$P(\hat{p}_i^R > \gamma | \alpha_i < \gamma, MA, \Pi)$
$P(\hat{p}_i^W > \gamma | \alpha_i < \gamma, \cdot W, \Pi)$
$P(\hat{p}_i^F > \gamma | \alpha_i < \gamma, F\cdot, \Pi)$

Figure 2: Induced probabilities for students to enter the math competition given their type and optimal strategy σ_i^*

Male students with an Asian last name have on average the largest induced probability to choose enter the competition. This is driven by the fact that they are most likely to make a Type-I error, and least likely to make a Type-II error. Female students with a Western last name have on average the smallest probability to enter the competition, since they are most likely to make a Type-II error and least likely to make a Type-I error. Male students with a Western last name and female students with an Asian last name can decrease the likelihood of making both types of error. This creates the following order on the induced probabilities of entering the competition.

$$\Phi(\alpha_i, MA, \sigma_i^*, \Pi) > \Phi(\alpha, MW, \sigma_i^*, \Pi) = \Phi(\alpha_i, FA, \sigma_i^*, \Pi) > \Phi(\alpha_i, FW, \sigma_i^*, \Pi)$$

At the same time, female students with a Western last name cannot boost up their beliefs with the use of social identity cues. Therefore, they will choose to enter the competition for higher realizations of $\hat{\alpha}_i$. Because these noisy perceptions are unbiased, conditional on entering the competition, they will have on average a higher success rate. The opposite reasoning applies to male students with an Asian last name. This creates the following order on success rates conditional on entering the competition.

$$E(\alpha_i | a_i = C, MA) < E(\alpha_i | a_i = C, \Theta_i \in \{MW, FA\}) < E(\alpha_i | a_i = C, FW)$$

More generally, let $\theta^{k\prime} \in \{0,1\}$ be the complement of θ^k. Let $\tilde{t}_{\theta^k}$ be the *one-sided* social type that has $\theta_i^k = \theta^k$ for all $k \in \{A,B\}$, while $\tilde{t}_{\theta^{k\prime}}$ is the *one-sided* social type that has $\theta_i^k = \theta^{k\prime}$ for all $k \in \{A,B\}$. Let T_{mixed} the set of *mixed* social types.

COROLLARY 1: *Let $\pi_{\theta^k} > \overline{\pi}_{\theta^k}$ for all $k \in \{A,B\}$. We have a* ***type-specific population effect****, such that $\Phi(\alpha_i, \tilde{t}_{\theta^k}, \sigma_i^*, \Pi) > \Phi(\alpha_i, t \in T_{mixed}, \sigma_i^*, \Pi) > \Phi(\alpha_i, \tilde{t}_{\theta^{k\prime}}, \sigma_i^*, \Pi)$, and a* ***type-specific selection effect****, such that $E(\alpha_i|a_i = C, \tilde{t}_{\theta^k}) < E(\alpha_i|a_i = C, t \in T_{mixed}) < E(\alpha_i|a_i = C, \tilde{t}_{\theta^{k\prime}})$.*

Differences in options agents have to bias the noisy perception $\hat{\alpha}_i$ create both a difference in the propensity with which agents with a certain social type undertake the *Competence-Driven* task and a difference in mean-competence conditional on undertaking this task. These patterns arise, even when the two traits of which a social type consists are independent from each other, and the differences between the respective social identity cues and their benchmark are induced by independent sources

4.2.3 Persistence

Whether these population and selection effects in Corollary 1 create persistent differences in choice behavior across social types depends on whether they shrink or increase the differences between the cues and their benchmark.

Let $\pi^*_{\theta^k} \in [0,1]$ be the value of π_{θ^k} in equilibrium. I characterize the possible equilibrium outcomes as follows[11].

DEFINITION 4 (Population Equilibrium): *In a '****Symmetric Equilibrium****' the allocation of agents over tasks is independent of their social type, and we have a fixed point such that $\overline{\Pi} = \tilde{\Pi}(\sigma^*, \overline{\Pi})$, where $\overline{\Pi}$ is such that $\pi^*_{\theta^k} = \overline{\pi}_{\theta^k}$ for $\theta^k \in \{0,1\}$ and $k \in \{A,B\}$. In an '****Asymmetric Equilibrium****' the allocation of agents over tasks is different for agents with a different social type, and we have a fixed point such that $\Pi = \tilde{\Pi}(\sigma^*, \Pi)$, where Π is such that $\pi^*_{\theta^k} \neq \overline{\pi}_{\theta^k}$ for $\theta^k \in \{0,1\}$ and $k \in \{A,B\}$.*

[11] The one-dimensional Asymmetric Equilibrium is considered in Liqui Lung (2022)

Furthermore, Definition 4 defines when a *Symmetric Equilibrium* becomes *unstable.*

DEFINITION 5 (Stability): *A Symmetric Equilibrium is* ***stable*** *when, for any value* $\delta > 0$*,* $\tilde{\pi}_{\theta^k}(\sigma^*, \overline{\pi}_{\theta^k} + \delta) < \overline{\pi}_{\theta^k} + \delta$ *for all* $k \in \{A, B\}$*. A Symmetric Equilibrium is* ***unstable*** *when there exists a value* $\delta > 0$*, such that* $\tilde{\pi}_{\theta^k}(\sigma^*, \overline{\pi}_{\theta^k} + \delta) > \overline{\pi}_{\theta^k} + \delta$ *for at least one dimension* $k \in \{A, B\}$*.*

Example - Assume students have an extreme response function such that for $\sigma_i\{\theta^A, \theta^B\}$,

$$\Phi(\alpha_i, \Theta_i, \sigma_i^*, \Pi) = \begin{cases} 1 & \text{when } \pi_{\sigma_i^*} > \overline{\pi}_{\sigma_i^*} \\ \Phi(\alpha_i, \theta_i, R, \Pi) & \text{when } \pi_{\sigma_i^*} = \overline{\pi}_{\sigma_i^*} \\ 0 & \text{when } \pi_{\sigma_i^*} < \overline{\pi}_{\sigma_i^*} \end{cases}$$

We can show a *Symmetric Equilibrium* always exists. Take a social context Π such that $\pi_{\sigma_i^*} = \overline{\pi}_{\sigma_i^*}$ for all i, meaning no trait is relatively over- or underrepresented in the data. This implies the strategies $\sigma \in \{\theta^A, \theta^B, R\}$ are equivalent for all social types Θ. Therefore, there will be no differences in the induced choice behavior across social types, and $\tilde{\pi}_{\theta^k}(\sigma^*, \overline{\pi}_{\theta^k}) = \overline{\pi}_{\theta_k}$ for all $k \in \{A, B\}$ and $\theta^k \in \{0, 1\}$.

Now, assume an exogenous shock δ to Π, such that $\pi_M = \overline{\pi}_M + \delta$ and $\pi_A = \overline{\pi}_A + \delta$. That is, now, there are slightly more male students and students with an Asian last name among those that qualify for the international final. With the extreme response function, all *talented* male students and students with an Asian in the next generation will enter the competition with probability one, while all *not talented* female students and students with a Western last name will choose the generic task. Consequently, $\tilde{\pi}_M(\sigma, \overline{\pi}_M + \delta) > \overline{\pi}_M + \delta$ and $\tilde{\pi}_A(\sigma, \overline{\pi}_A + \delta) > \overline{\pi}_A + \delta$ and hence, the *Symmetric Equilibrium* becomes unstable.

Using this extreme response function, we can furthermore show that the induced social identity cues $\tilde{\pi}_{\theta^k}(\sigma, \Pi)$ are bounded from above. Let the induced number of individuals successful in the competition with trait θ^k be denoted by,

$$S_{\theta^k} = p_{\theta^k} \int \alpha \Phi(\alpha, \Theta : \theta_i^k = \theta_k, \sigma^*, \Pi) d\alpha$$

Then,

$$S_M = p_M \int_{\alpha > \gamma} \alpha f(\alpha) d\alpha + p^M p^A \int_{\alpha < \gamma} \int_{\hat{\alpha} > \gamma} \alpha g_\alpha(\hat{\alpha}) f(\alpha) d\alpha d\hat{\alpha} \tag{1}$$

is the total number of male students successful in the competition, and,

$$S_F = p_F p_W \int_{\alpha > \gamma} \int_{\hat{\alpha} > \gamma} \alpha g_\alpha(\hat{\alpha}) f(\alpha) d\alpha d\hat{\alpha} + p_F p_A \int_{\alpha > \gamma} \alpha f(\alpha) d\alpha \tag{2}$$

is total number of female students. Consequently, we can write

$$\frac{\tilde{\pi}_M(\sigma^*, \Pi)}{\tilde{\pi}_F(\sigma^*, \Pi)} \leq \frac{S_M}{S_F} \tag{3}$$

We can obtain a similar equation for the dimension of last names. Therefore, for any response function, $\frac{S_M}{S_F}$ provides an upper bound on $\frac{\tilde{\pi}_M(\sigma^*,\Pi)}{\tilde{\pi}_F(\sigma^*,\Pi)}$. This is sufficient to show that, when the Symmetric Equilibrium becomes unstable in two dimensions, we move towards an Asymmetric Equilibrium of degree 2.

Finally, the induced number of successful individuals of a certain social type Θ, denoted by $S_\Theta = p_\Theta \int \alpha \Phi(\alpha, \Theta, \sigma^*, \Pi) d\alpha$, can be ordered in terms of social type, such that,

$$\begin{aligned}
S_{MA} &= p_M p_A \int_{\alpha > \gamma} \alpha f(\alpha) d\alpha + p^M p^A \int_{\alpha < \gamma} \int_{\hat{\alpha} > \gamma} \alpha g_\alpha(\hat{\alpha}) f(\alpha) d\alpha d\hat{\alpha} \\
S_{FA} &= p_F p_A \int_{\alpha > \gamma} \alpha f(\alpha) d\alpha \\
S_{MW} &= p_M p_W \int_{\alpha > \gamma} \alpha f(\alpha) d\alpha \\
S_{FW} &= p_F p_W \int_{\alpha > \gamma} \int_{\hat{\alpha} > \gamma} \alpha g_\alpha(\hat{\alpha}) f(\alpha) d\alpha d\hat{\alpha}
\end{aligned}$$

Hence, in an Asymmetric Equilibrium, male students with an Asian last name will be most overrepresented among those successful, due to the fact that the *not talented* students with this social type cannot decrease the likelihood of making a Type-II error. Similarly, female students with a Western last name will be most underrepresented, due to the fact that the *talented* students with this social type cannot decrease the likelihood of making a Type-I error. Proposition 3 generalized these results.

PROPOSITION 3: *If a* ***Symmetric Equilibrium*** *is unstable in both dimensions* $k \in \{A, B\}$*, then it co-exists with a stable* ***Asymmetric Equilibrium of Degree 2****. Assume WLOG that in any* ***Asymmetric Equilibrium of Degree 2*** $\pi_{\theta^k} > \overline{\pi}_{\theta^k}$ *for* $k \in \{A, B\}$*. Then, with* $S_\Theta = p_\Theta \int \alpha\Phi(\alpha, \Theta, \sigma^*, \Pi)d\alpha$*, the order on* S_Θ *must be such that,*

$$S_{\tilde{t}_{\theta^k}} > S_{T_{mixed}} > S_{\tilde{t}_{\theta^{k'}}}$$

This proposition shows how shocks in the representation of independent dimensions of a social type in social context interact and can induce persistent differences in choice behavior and success rates across social types. Consequently, differences in the options to bias decision making across social types can make multidimensional asymmetries in social context a self-fulfilling prophecy.

The order on the number of successful individuals across social types in an Asymmetric Equilibrium is consistent with what we find in the data. The 'Leaders and Daughters Global Survey 2019' and Pogrebna et al. (2024) shows women's career trajectories flatten mid-career compared to men's, where this downward trend is strongest for women belonging to ethnic minorities. Similarly, Gupta (2019) and Charleston et al. (2014) show how women belonging to underrepresented castes, respectively African American women are disproportionally underrepresented in STEM subjects. Furthermore, the implied order on the number of successful individuals across social types in this model is the similar to what is we find through other mechanisms studied in multidimensional contexts, such as discrimination and unequal access to public goods (e.g. Crenshaw (1991), Yuval-Davis (2015)). In Liqui Lung (2022), I integrate such direct effects of social context on utility in the model, and indeed show how the different mechanisms reinforce each other in creating persistent differences in choice behavior across a priori identical agents belonging to different social groups.

Finally, with a model of belief formation in mind, each strategy can be interpreted as a choice of model about the relationship between traits and chances of success in the *Competence-Driven* task. Another interpretation of Proposition 1 is that agents adopt the model that 'works best for them' in terms of expected utility. Moreover, following Proposition 3, equilibrium beliefs are consistent with observational feedback conditional on the model of the world an agent adopts. That is, a *talented* female agent only adopts a model of the world in which she believes gender and outcomes are correlated when female agents are socially more successful. When this is the case, adopting this model will ensure she will make on average less mistakes in decision making and hence, observe on average more successful outcomes. As she is a female agent herself, this feedback confirms her belief that female agents are more successful. Hence, even when the model agents end up adopting according to their fitness criterion is misspecified, the observational feedback provides a narrative that is consistent with this misspecified model.

4.2.4 Existence and the Degree of Asymmetry

The existence of an Asymmetric Equilibrium of Degree 2 is not trivial. Furthermore, the degree to which differences in choice behavior in an *Asymmetric Equilibrium* contribute to the persistence of inequalities across social groups depends on how large the differences between the cues π^* and their benchmarks $\overline{\pi}$ are in such an equilibrium. In the following, I give an intuitive overview of the factors that influence both existence and the degree of asymmetry.

Response Function and Outside Option - For an Asymmetric Equilibrium to exist, a Symmetric Equilibrium needs to become unstable as the result of a perturbation. Whether a Symmetric Equilibrium is unstable depends on γ and the properties of the response function $\eta(\pi_{\theta^k}, \overline{\pi}_{\theta^k})$. First, an exogenous shock δ to Π in a '*Symmetric Equilibrium*' must have a sufficiently large effect on choice behavior of agents at the individual level. Specifically, the induced change in the threshold $\gamma^{\sigma_i^*}$ must be large enough. This change depends both on the derivative of the response function $\eta(\pi_{\theta^k}, \overline{\pi}_{\theta^k})$ at the '*Symmetric Equilibrium*', and, because of the linearity of $\gamma^{\sigma_i^*}$ in γ, this change is multiplicative in γ. Secondly, a perturbation δ must have a sufficiently large effect on the outcomes at the aggregate level. This is captured by the elasticity of the total number of successful people in the *Competence-Driven* task in γ. The absolute value of this elasticity is increasing in γ, since the more attractive the outside option, the lower the number of agents that tries the *Competence-Driven* task. Moreover, the higher γ, the higher the success rate of agents that choose this task. Consequently, the effect of a change in behavior on the induced social context $\tilde{\Pi}(\sigma, \Pi)$ is increasing in γ. Finally, the stronger agents respond to differences between π and $\overline{\pi}$, the larger the difference π^* and $\overline{\pi}$ in an Asymmetric equilibrium.

COROLLARY 2: *Take two response functions $\hat{\eta}$ and η, such that $\hat{\eta}(\pi, \overline{\pi}) > \eta(\pi, \overline{\pi})$ for all $\pi > \overline{\pi}$. Assume WLOG that an 'Asymmetric Equilibrium' exists in which $\pi > \overline{\pi}$. Let π^*_η be the equilibrium value of π given a response function η. Then, $\pi^*_{\hat{\eta}} > \pi^*_\eta$.*

Distribution of Social Types - A second factor that contributes to the existence of an Asymmetric Equilibrium and the degree of asymmetry in such an equilibrium is the fraction of agents with a *mixed* social type relative to the fraction of agents with a *one-sided* social type in the population. Specifically, the larger the fraction of agents with a *one-sided* social type, easier a Symmetric Equilibrium becomes unstable and the larger the differences between $\pi^*_{\theta^k}$ and $\overline{\pi}_{\theta^k}$ in equilibrium for $k \in \{A, B\}$. The intuition for this result is as follows.

Example - The population and selection effects are driven by the fact that agents with a different value of a particular trait, e.g. men and women, make on average a different type of mistake. As agents with a *mixed* social type can minimize the likelihood of making both types of mistakes, this asymmetry is predominantly driven by agents with a *one-sided* social type. We can illustrate this intuition with Equations 1 and 2. When we increase the fraction of agents with a *mixed* social type, we increase p_{MW} and p_{FA}. As all *not talented* male students with a Western last name choose the generic task, this means the number of *not talented* male students mistakenly entering the competition goes down. This induces a decrease in S_m. On the other hand, an increase in the fraction of female students with an Asian last name leads to an increase in S_F, as all *talented* students with these traits will surely enter the competition. As a result, an increase in the fraction of agents with a *mixed* social type leads to a decrease in the strength of the population effects and lowers the upper bound $\frac{S_M}{S_F}$ on $\frac{\tilde{\pi}_M(\sigma^*,\Pi)}{\tilde{\pi}_F(\sigma^*,\Pi)}$.

Dimensionality of Social Type - The strength of the population effects along the lines of any single trait θ^k is a function of the number of traits k along which there is asymmetry in social context. Specifically, we can show that, as the degree k of a social type increases, the strength of the population effects in any particular dimension decreases. The intuition behind this result is as follows.

Example - Assume one-dimensional social types based on the *Gender*-dimension. With the extreme response function, we get,

$$S_M = p_M \int_{\alpha>\gamma} \alpha f(\alpha) d\alpha + p_M \int_{\alpha<\gamma} \int_{\hat{\alpha}>\gamma} \alpha g_\alpha(\hat{\alpha}) f(\alpha) d\alpha d\hat{\alpha} \tag{4}$$

while the number of successful female students will be,

$$S_F = p_F \int_{\alpha>\gamma} \int_{\hat{\alpha}>\gamma} \alpha g_\alpha(\hat{\alpha}) f(\alpha) d\alpha d\hat{\alpha} \tag{5}$$

Now we move back to the two-dimensional case, as presented in Equations 1 and 2. The difference between Equations 1 and 4 is that, because of the introduction of the *Name*-dimension, S_M decreases through the fraction of *not talented* male students with a Western last name that can now use the *Name*-dimension to correct for a Type-II error. Similarly, the difference between Equations 2 and 5 is that S_F increases, because *talented* female students with an Asian last name can now use the *Name*-dimension to correct for a Type-I error. As we continue to increase the number of dimensions k of the social type, the fraction of the population with a *mixed* social type increases. This will decrease the strength of the population effects in any particular dimension k, making it more difficult for a Symmetric Equilibrium to become unstable. Furthermore, it lowers the upper bound on $\frac{\tilde{\pi}_M(\sigma^*,\Pi)}{\tilde{\pi}_F(\sigma^*,\Pi)}$, decreasing the asymmetry in choice behavior that can persist when an Asymmetric Equilibrium exists.

Aggregate Expected Utility and Multidimensionality - When we aggregate expected utility over the population, any *Asymmetric Equilibrium* is a Pareto improvement over a *Symmetric Equilibrium*, as, in an *Asymmetric Equilibrium*, only those agents that can improve decision making on average with the *social identity cues* change their behavior. The agents that cannot use social context to improve decision making are not made worse off. This is no longer necessarily true when beliefs have a direct effect on the probability of success, through for example confidence (Compte and Postlewaite, 2004), or when social context has a direct effect on someone's chances of success through some form of discrimination or stereotype threat (Steele, 2010). An '*Asymmetric Equilibrium*' can also become suboptimal when agents make systematic errors in learning their optimal strategy, when they do not correctly compute the long-term pay-offs of choosing a *Competence-Driven* task, or when the strategy *Repress* becomes costly. Most importantly, we may want to avoid an '*Asymmetric Equilibrium*' when it contributes to the persistence of inequalities across social groups by reinforcing harmful stereotypes, social norms, or statistical discrimination.

Whether an *Asymmetric Equilibrium of Degree k* increases aggregate utility over a *Asymmetric Equilibrium of Degree k-1* depends on the following trade-off. On the one hand, when agents have access to more social identity cues, the fraction of agents with a *mixed* social type increases. This means that, on average, the set of agents that can potentially improve decision making becomes larger, which has a positive effect on aggregate utility. On the other hand, the degree of asymmetry along the lines of a particular trait θ^k is driven by the asymmetry in the types of error subgroups can potentially correct for. In an *Asymmetric Equilibrium of Degree 1*, this asymmetry is driven by the behavior of all agents. In a *Asymmetric Equilibrium of Degree* $k > 1$, this asymmetry is only driven by the agents with a *one-sided* social type. We therefore observe smaller deviations of $\pi^*_{\theta^k}$ from $\overline{\pi}_{\theta^k}$ as k increases. These smaller deviations lead to smaller effects of the social identity cues on decision making, and decrease the potential of agents to improve decision making. As we keep on adding dimensions to the social type, this negative effect on aggregate utility starts to overtake. The deviations of π_{θ^k} from $\overline{\pi}_{\theta^k}$ decrease, until the share of agents in the population with a *one-sided* social type is too small to make a *Symmetric Equilibrium* unstable.

4.3 Adding Two-Dimensional Social Identity Cues

To simplify the exposition of the model, I assumed agents could only use *one-dimensional* social identity cues. In this section, I show what the effects are of adding the option of using the *two-dimensional* social identity cues π_t on both individual and aggregate choice behavior. The introduction of this option changes the strategy set to $\sigma_i \in \{\theta^A, \theta^B, \Theta, R\}$, where Θ refers to the strategy in which agents use the *two-dimensional* social identity cue derived from their full social type Θ_i. When $\sigma_i = \Theta$, the corresponding response function can be defined as,

$$\eta_\Theta = \eta(\pi_{\Theta_i}, \overline{\pi}_{\Theta_i})$$

The function η_Θ could in principal be different from η_{θ^k}. People could for example react stronger to *two-dimensional* social identity cues than *one-dimensional* social identity cues, because they identify stronger with people that have their entire social type Θ_i in common, instead of only one trait θ_i^k. The strategy $\sigma_i = \Theta$ results in a belief $\hat{p}^\Theta = \eta_\Theta \hat{\alpha}_i$ or a threshold $\gamma_i^\Theta = \frac{\gamma}{\eta_\Theta}$. We can similarly redefine

$$\overline{\overline{\eta}}_\Theta = \max(\max_k(\eta_{\theta^k}), \eta_\Theta) \text{ and } \underline{\underline{\eta}}_\Theta = \min(\min_k(\eta_{\theta^k}), \eta_\Theta)$$

Similarly, we define

$$\overline{\overline{\kappa}}_\Theta = \underset{\Theta,k}{\text{argmax}}(\eta_{\theta^k}, \eta_\Theta) \text{ and } \underline{\underline{\kappa}}_\Theta = \underset{\Theta,k}{\text{argmin}}(\eta_{\theta^k}, \eta_\Theta)$$

Hence, $\overline{\overline{\eta}}_\Theta \geq \overline{\eta}_\Theta$, while $\underline{\underline{\eta}}_\Theta \leq \underline{\eta}_\Theta$. Adding the strategy Θ provides the following Corollary to Proposition 1.

COROLLARY 3: *The individually optimal strategies σ_i^* given an agent's type $\{\alpha_i, \Theta_i\}$ with $\sigma_i \in \{\theta^A, \theta^B, \Theta, R\}$ and a social context Π are the following:*

	$\alpha_i > \gamma$	$\alpha_i < \gamma$
$\overline{\overline{\eta}}_\Theta > 1$	$\overline{\overline{\kappa}}_\Theta$	R
$\underline{\underline{\eta}}_\Theta < 1$	R	$\underline{\underline{\kappa}}_\Theta$

Corollary 3 shows agents will only use the option $\sigma_i = \Theta$, when η_Θ provides a stronger bias in the direction of their welfare-maximizing task than η_{θ^k} for any k.

The extra strategy affects choice behavior at the aggregate level in the following ways. Again, let $\tilde{t}_{\theta^k}$ be the *one-sided* social type that has $\theta_i^k = \theta^k$ for all $k \in \{A, B\}$, while $\tilde{t}_{\theta^{k\prime}}$ is the *one-sided* social type that has $\theta_i^k = \theta^{k\prime}$ for all $k \in \{A, B\}$. Let T_{mixed} the set of *mixed* social types. Assume $\pi_{\theta^k} > \overline{\pi}_{\theta^k}$ for all $k \in \{A, B\}$. First, when $\eta_{\tilde{t}_{\theta'}} < \eta_{\theta^{k\prime}}$ for $k \in \{A, B\}$ and $\eta_{\tilde{t}_{\theta^k}} > \eta_{\theta^k}$ for $k \in \{A, B\}$, the introduction of the strategy $\sigma_i = \Theta$ increases the *type-specific population and selection effects.*

Secondly, η_Θ may induce a stronger bias for a similar pair $(\pi, \overline{\pi})$ than η_{θ^k}. Tt may then become optimal for agents with a *mixed* social type to choose $\sigma_i = \Theta$. This will create an asymmetry in the potential to improve decision making across agents with a mixed social type. Definition 6 helps analyze what happens at the aggregate level.

DEFINITION 6: *An observable trait θ^k is dominant when $\pi_\Theta > \overline{\pi}_\Theta$ for all $t : \theta_i^k = \theta^k$, while $\pi_\Theta < \overline{\pi}_\Theta$ for all $t : \theta_i^k = \theta^{k\prime}$.*

In other words, we call an observable trait *dominant*, when agents with a *mixed* and *one-sided* social type with one realization of this trait, e.g. all male students, are overrepresented, while agents with a *mixed* and *one-sided* social type with the other realization of this trait, e.g all female students, are underrepresented. For a trait that is not dominant, for one realization agents with a *one-sided* social type are overrepresented, e.g. male students with an Asian last name, while agents with a *mixed* social type are underrepresented, e.g. female students with an Asian last name, and for the other realization vice versa, e.g. male and female students with a Western last name. When agents with a *mixed* social type choose $\sigma_i = \Theta$, this increases the population and selection effects along the lines of the *dominant* observable trait, while it decreases the population and selection effects along the lines of the trait that is not dominant.

5 What fosters Persistent Identity-Driven Choices?

When we consider multidimensional social types, the list of traits one can think of can be infinite. Yet, as we saw in the previous section, differences in choice behavior can only persist along the lines of a limited number of traits simultaneously. Moreover, in reality, traits that drive choice behavior are often related to gender, ethnicity, age or social class. In this section, I show what insights the model can provide regarding what characteristics of a trait foster persistent differences in choice behavior.

5.1 The Ability to Self-Identify

People have a certain degree of control over their social type. Jia and Persson (2019) shows how the choice of ethnicity for children in ethnically mixed marriages in China is driven by the interaction between material benefits restricted to certain minorities, and existing social norms of following the father's identity. Qian and Nix (2015) shows how the rate at which black Americans were 'passing' as white was correlated with geographical relocation to communities with higher percentages of whites, and with better political and socioeconomic opportunities for whites relative to blacks. Cassan (2015) shows how the Punjab alienation of land act led to a movement of identity-manipulation. Furthermore, there are many flexible traits that are easy for people to adopt, think of fashion styles, belonging to certain subcultures or societies.

The model allows to study the ability to self-identify through enlarging the strategy set. This ability may be costly for some traits, like gender, while more flexible for others, like having a certain hair color. To guide intuition on how this affects the equilibrium results, consider an extreme case in which agents can freely choose their trait θ_i^k. This means we effectively endogenize this dimension of the social type. When agents can pick and choose their social type as they like, they can decrease the likelihood of making both a Type-I and Type-II error. This is beneficial for agents at the individual level, as it will increase their expected utility. Yet, these benefits cannot be persistent. At the aggregate level, when agents can fully self-identify, the whole population behaves as an agent with a *mixed* social type. Consequently, an Asymmetry Equilibrium does not exist. As soon as agents benefit from any asymmetries along the lines of this trait at the individual level, these asymmetries will disappear in the next generation. As we restrict the ability of agents to self-identify, or make it more costly, we move from this extreme case towards the case in which the social type is fully exogenous. This increases the size of the population that has a *one-sided* social type. This makes it easier for an Asymmetric Equilibrium to exist and increases the degree of asymmetry that can possibly persist in equilibrium.

Hence, the more restricted agents are in their ability to self-identify, the stronger the population and selection effects in this dimension of the social type. This provides intuition for why traits having persistent effects on choice behavior are traits that are costly to manipulate, such as gender, race or social class, while traits that only have a short-lived effect on behavior, such as fashion styles, are easy to adopt. Even if such flexible traits may drive behavior for some time, the effects will eventually dissipate, guiding agents' focus back to less their flexible traits.

5.2 The Cost of Repressing

It may be costly to repress certain dimensions of the social type because of socially imposed constraints. Specifically, one's social identity is a composite view of the view one has of oneself as well as the views held by others about one's identity (Nagel, 1994). The views held by others may be guided by stereotypes, narratives or stigmatization. It can be difficult for agents to ignore the dimension of their social type that is made salient by such social constraints (Major and O'Brien, 2005). Furthermore, it can be difficult to identify with other social types, even if they have some traits in common. For example, Crenshaw (1991) describes how black women are not able to identify with white women nor black men, because their experience in society is so different.

The framework developed in this paper allows us to study the effects of social constraints through simple adjustments to the strategy set. To guide intuition, I again consider various extreme cases in which the cost of repressing is infinite and agents are not able to use certain identification strategies σ_i. In particular, I consider two types of strategy restrictions; The first type is such that agents are not able to ignore one dimension of their social type, while they are free to ignore the other dimension. I summarize constraints of this type under the name *Stigmatization.* The second type of restriction is such that agents can only use *two-dimensional* social identity cues in belief formation. Such *Type-Specific Social Identification* captures agents that cannot identify with other social types, despite having some traits in common.

5.2.1 Stigmatization

I consider two versions of *Stigmatization.* In the first version, an entire trait θ_i^k is stigmatized. I first analyze a *Two-Strategy Model*, in which agents can only use *one-dimensional* cues. Consequently, I introduce the *Three-Strategy Model*, in which agents can use both one-dimensional cues π_{θ^k} and two-dimensional cues π_t. In the second version of stigmatization, only one value of a trait is stigmatized, meaning that only those agents with that specific value of the trait cannot ignore the trait, e.g. women, while agents with another value can, e.g. men. I call this model the *Asymmetric Model.* I use the *Two-Strategy Model* to provide the main intuition and discuss the main take-aways from the full analysis below. The complete analysis of the *Three-Strategy Model* and the *Asymmetric Model* can be found in Appendix 2.

Two-Stategy Model - In a two-strategy model, agents only use *one-dimensional* social identity cues π_{θ^k} and cannot repress one of their traits θ_i^k. Assume gender is the stigmatized dimension of social identity. Hence, the strategy set is reduced to $\sigma_i \in \{Gender, R\}$. In this case, there is no difference in the potential to improve decision making between *mixed* and *one-sided* social types. Only students with *mixed* types are disadvantaged by stigmatization in this model, while students with *one-sided* types are not affected. Furthermore, students cannot use the dimension of their last name in decision making. Because the traits θ_i^k and ability types α_i are independently distributed over the population, there can be no differences in choice behavior between students with a Western and Asian last name. On the other hand, the differences in choice behavior between male and female students are now driven by the behavior of students with both one-sided social types and *mixed* social types. Therefore, stigmatization reinforces the population and selection effects in the dimension of gender.

The main take-aways from the full analysis are that, first, *Stigmatization* reinforces the population and selection effects in the dimension of the stigmatized trait, and decreases these effects in the dimension of the non-stigmatized trait. Secondly, *Stigmatization* mainly has a negative effect on the potential to improve decision making of agents with a *mixed* social type. When the stigmatized trait is not *dominant*, the availability of *two-dimensional* social identity cues can partially mitigate these negative effects, as agents with a *mixed* social type can substitute, at least to some extent, the *one-dimensional* cue they can no longer use with their *two-dimensional* cue π_{Θ}.

5.2.2 Type-Specific Social Identification

When agents can only focus on others with their social type Θ, this reduces the strategy set to $\sigma_i \in \{\Theta_i, R\}$. When we eliminate the option to use *one-dimensional* social identity cues, all social types only have the ability to decrease the likelihood of making one type of mistake. When gender is *dominant*, all male students can potentially correct for a Type-II error, while all female students can potentially correct for a Type-I error. This induces population and selection effects in the dimension of gender that are driven by the behavior of agents with both *one-sided* and *mixed* social types. On the other hand, in the dimension that is not *dominant*, male students with an Asian last name can potentially correct for a Type-II error, while female students with an Asian last name can potentially correct for a Type-I error. For male and female agents with a Western last name, this is exactly the opposite. Hence, there will only be asymmetry along the lines of the last names, when $\pi_{MA} \neq \pi_{MW}$ and $\pi_{FW} \neq \pi_{FA}$. The exact opposite happens when the last name is the *dominant* trait. Therefore, compared to the benchmark case, *Type-Specific Social Identification* increases the population and selection effects along the lines of the *dominant* dimension of the social type, while decreases asymmetry along the lines of the trait that is not dominant.

5.3 Correlation

Correlated traits - When two socially more successful traits θ^A and θ^B are positively correlated, there are relatively more agents with a *one-sided* social type than a *mixed* social type in the population. This is the type of correlation we often observe. For example, belonging to an underrepresented ethnic minority is often correlated with belonging to a lower income group, while the opposite applies to overrepresented minorities (e.g. Asian origin) and higher income groups. Similarly, belonging to certain social groups can be correlated with having certain physical aspects, or it can make some attributes more easily accessible. This type of correlation has two effects. First, because it increases the fraction of agents with *one-sided* social types in the population, it increases the strength of the population and selection effects. Secondly, a shock δ to social context in one dimension of the social type will simultaneously affect social context in the correlated dimension. This will further reinforce the resulting population and selection effects. This type of correlation makes it therefore easier for an Asymmetric Equilibrium to exist. On the other hand, when the correlation is negative, this increases the fraction of agents with a *mixed* social type, which decreases the degree of asymmetry in equilibrium.

Correlation with Ability - To isolate the mechanism through which the results are obtained from other mechanisms, such as social learning, I assumed social types Θ_i are uncorrelated with ability types α_i. Yet, when there is a correlation between a trait θ^k and ability, this has a direct and indirect effects on the induced population and selection effects. Consider the case in which a trait θ^k, e.g. height, is correlated with ability α_i, playing basketball. That is, the fraction of agents in the population that are tall and *talented* is larger than the fraction of agents that are small and *talented*. Therefore, the population and selection effects in a social context in which tall people are overrepresented among those successful will be stronger than these effects in a context in which small people are overrepresented. As a consequence, it is easier for an Asymmetric Equilibrium to exist of the former type than the latter.

Secondly, if one socially more successful trait θ^A, e.g. height, is correlated with both ability and another socially more successful trait θ^B, e.g. belonging to a certain ethnic group, this increases both the fraction of *talented* agents in the population with a socially more successful *one-sided* social type and the fraction of *not talented* agents with a socially less successful *one-sided* social type. This reinforces the population and selection effects and increases the degree of asymmetry observed in an Asymmetric Equilibrium in which tall people from this ethnic group are relatively overrepresented among those successful in basketball. Contrary to the case in which there was no correlation with ability, the social type is now a meaningful instrument for agents.

6 Policy Implications: Pitfalls and Opportunities

6.1 The Pitfalls

When it comes to designing policy, a multidimensional view is not only helpful, but crucial. Recent literature challenges the effectiveness of the traditional one-dimensional policy measures employed to achieve social diversity. Such approaches, like those centered on gender alone, can not only have limited impact, but can induce unintended spillover effects on the representation of other social groups (See for example Cassan and Vandewalle (2021), Beaman et al. (2012), Hughes (2011), Karekurve-Ramachandra and Lee (2020), Folke et al. (2015) and Tan (2014)). When we analyze data on choice behavior and outcomes with a *one-dimensional view* on social types, there are important aspects we would overlook. In this section, I show how these insights could explain the pitfalls we encounter in the literature.

Externalities - The patterns in aggregate choice behavior in Corollary 1 induce externalities in choice behavior and success rates along the lines of individual traits as the result of changes in social context. For example, when we eliminate the differences in representation among male and female students in a math contest, we would take away the option of not talented female Asian students to bias beliefs downwards, and of talented male Western students to bias beliefs upwards. This leads to an increase in Asian students among those that participate and succeed, and a decrease in Western students. This is consistent with what we find in the data. Results from the ATfoundation Math Prize for girls shows the list of winners is almost entirely composed of women of Asian origin.[12] These externalities can explain the unwanted spillover effects of gender quota we encounter in the literature. For example, Cassan and Vandewalle (2021) shows a quota to increase the participation of women in leadership positions in India reinforced the overrepresentation of people belonging to certain castes. The

[12] See https://mathprize.atfoundation.org/experience/past-events for more information.

model would predict this being the result of not talented women belonging to already overrepresented castes now being less able to bias beliefs downwards with social identity cues related to their gender. They are therefore more likely to enter leadership positions, further strengthening the overrepresentation of these castes.

One-Dimensional Population and Selection Effects - When we look at Table 1 with a *one-dimensional view* along the lines of *Gender*, on average, male and female students have the same potential to correct for the mistakes. There are nevertheless more male students that can minimize the likelihood of making a Type-II error, while there are more female students that can minimize the likelihood of making a Type-I error. Figure 2 shows how, consequently, male students have on average a larger probability to enter the competition than female students. At the same time, female students have on average a higher success rate than male students conditional on entering the competition. We have therefore *one-dimensional* population and selection effects. Without a *multidimensional view*, we would nevertheless fail to understand these effects are predominantly driven by the behavior of students with a *one-sided* social type, caused by both the inequality in expected utility between agents with a *mixed* and *one-sided* social type, and the asymmetry in the type of error agents with a different *one-sided* type can potentially correct for. Crenshaw (1991) discusses how gender quota affect white women disproportionately. This paper suggests we could improve the effectiveness of this policy and reduce spillover effects by making sure a quota affects all women proportionately, or even target minority women directly.

Stigma and Correlation - Restrictions on the strategy set induced by stigma or correlation between traits have externalities on choice behavior across different groups. When being female is stigmatized, this affects reactions to changes in social context along the lines of ethnicity. Moreover, when African American women are stigmatized and poorly represented in a given task, this will undermine the prevalence of women in general in that task, and, hence, negatively affect the option to bias beliefs upwards for all talented women. Not being aware of the restrictions agents face and the externalities they induce affects the effectiveness of policy. For example, when we want to avoid people to focus on ethnicity in a particular choice setting, for example through no longer making this information available, then this policy will have little to no effect when ethnicity is correlated with another variable on which agents do have information, such as social class. A similar problem arises when social class is stigmatized and we reduce asymmetries along the lines of ethnicity with a quota.

Welfare - Assume in university admissions we impose a measure that leads to an equal representation of students from different high school districts. Let us analyze the consequences of such a measure on choice behavior in the next generation using the simplified framework developed in this paper. As a result of the policy, students in the next generation will no longer focus on the role of the district they come from. Instead, they may now only be able to focus on gender. Consequently, we have taken away the option for agents with a *mixed* social type to use the high school district dimension to improve their decision making. These agents will now be more likely to make mistakes in decision making. Indeed, although the policy measure will result in an equal representation of students from different districts, it is obtained through an increase in *not talented* male students that apply and a decrease of *talented* female students. As result of the policy, we may therefore have decreased aggregate welfare.[13]

6.2 Opportunities

This paper shows how agents can use social identity cues to improve decision making on average. However, under certain conditions, these individual benefits come at a cost at the aggregate level, where the resulting behavior may contribute to the persistence of inequalities and harmful stereotypes. Optimally, we would like to eliminate the per-

[13] For simplicity, I did not consider the option that students know about the quota. This assumption is realistic for this particular example, where prospective students are usually not aware of the particulars of the admission procedure. If agents would be aware of a quota, this could affect the results in two ways. On the one hand, agents may believe social context is now less relevant in forming a belief about their own ability. This would flatten the functions η_{σ_i}. On the other hand, agents targeted by the quota may shift their beliefs about their chances of success systematically upwards, while agents not targeted by the quota shift their beliefs systematically downwards. This would further decrease asymmetries between the targeted and non-targeted subgroups.

sistent effects at the aggregate level, while maintaining the benefits at the individual level. A multidimensional view on social identity provides opportunities to achieve this. On the one hand, with could eliminate asymmetries in social context along the lines of traits that induce persistent effects with the use of well-designed multidimensional quota as discussed in the previous section. On the other hand, we can use informational policy to nudge people to focus on different dimensions of their social type or on different statistics, such that we avoid persistent negative effects at the aggregate level.

By making data available on certain traits and not others, we can guide attention to traits that are accessible, easy to manipulate, and with a low cost to repress. We could further enable agents to improve decision making when we direct focus to traits that are correlated with ability. This has not only a larger effect on behavior at the aggregate level, but also provides benefits in an informational sense. Secondly, by influencing the statistics agents calculate and process, we can eliminate the persistence of identity-driven choices. In particular, in Liqui Lung (2022), I show how social identity cues cannot have persistent effects on choice behavior when agents focus on within-group success rates. Third, we could influence agents' reaction to social context through their response function η. The more attention a certain trait receives through for example the media, the stronger agents will react to deviations of the social identity cues from their benchmarks. Finally, the reaction of agents to social context also depends on the benchmark they use. In Liqui Lung (2022), I show how a Symmetric Equilibrium no longer exists when agents have misspecified beliefs about their benchmark. Providing data and information regarding the correct benchmark is therefore another way in which we could avoid persistent asymmetries in choice behavior.

7 Conclusion

This paper provides a framework to analyze how the multidimensionality of social identity affects choice behavior at both the individual and aggregate level. I show how informationally irrelevant data about others can help agents improve decision making on average when they are subject to noise in decision making that they have limited tools to correct for. Agents can use social context to mechanically bias beliefs in a direction implied by their social type, and adopt the bias that best increases the likelihood they undertake their welfare-maximizing task.

The key insight is that, although agents have the same set of social identification strategies, social context translates this set of strategies into options to bias that are different for agents with different social types. It is therefore not the representation of a social group per se that affects behavior in this model, but rather the potential to improve decision making social context induces for each social type. Differences in this potential can induce persistent asymmetries in choice behavior across a priori identical individuals with different social types. The first message of the paper is therefore that, when we want to address identity-driven choice behavior, taking care of discrimination and initial skill differences is not enough.

The model allows to integrate the effects of stigma, identity manipulation and self-identification through changes in the set of social identification strategies. The insights obtained shed light on why we always talk about gender, ethnicity, social class and traits highly correlated with those, when it comes to occupational and educational choices. Finally, the equilibrium results shed light on the different externalities across traits that arise as a response to changes in social context. Knowledge of these externalities is key for the design of effective policy, and the second message of the paper is that a multidimensional approach to address the underrepresentation of e.g. women and minorities is crucial for effective policy design.

To conclude, social influences have a function in choice behavior and this paper aims to make a step towards opening this black box. The insights point towards various directions for future research. First, I assume homogeneity in the way agents observe and process information. Social networks may nevertheless influence an agent's perception of the social environment. This could create heterogeneous cues that may be correlated with traits such as income, neighborhood or education. Secondly, I assume agents are perfectly able to learn their optimal strategies. Social context could nevertheless influence this learning process, through for example discrimination, social pressures or stereotype threat. This could induce learning traps that could be asymmetric across social groups. A deeper understanding of these issues would allow us to better make the step from the theoretical framework to the real world.

References

G.A. Akerlof and R.E. Kranton. Economics and identity. *Quarterly Journal of Economics*, 115(3):715–753, 2000.

R. Akerlof. We thinking and its consequences. *American Economic Review*, 106(5): 415–419, 2016a.

R. Akerlof. Value formation: The role of esteem. *Games and Economic Behavior*, 102: 1–19, 2016b.

M.R. Banaji and A.G. Greenwald. *Blindspot*. Random House Publishing Group, 2016.

L. Beaman, E. Duflo, R. Pande, and P. Topalova. Female leadership raises aspirations and educational attainment for girls: a policy experiment in india. *Science*, 335 (6068):582586, 2012.

R. Benabou and J. Tirole. Incentives and prosocial behavior. *American Economic Review*, 96(5):1652–1678, 2006.

R. Benabou and J. Tirole. Identity, morals and taboos: Beliefs as assets. *The Quarterly Journal of Economics*, 126(2):805–855, 05 2011.

P. Bordalo, K. Coffman, N. Gennaioli, and A. Shleifer. Beliefs about gender. *American Economic Review*, 109(3):739–773, 2019.

M.K. Brunnermeier and J.A. Parker. Optimal expectations. *The American Economic Review*, 95(4):1092–1118, 2005.

Jean-Paul Carvalho, B. Pradelski, and C. Williams. Affirmative action with multidimensional identities. Working paper, 2023.

J.P. Carvalho and B. Pradelski. Identity and underrepresenation: Interactions between race and gender. *Journal of Public Economics*, 216:104764, 2022.

G. Cassan. Identity-based policies and identity manipulation: Evidence from colonial punjab. *American Economic Journal: Economic Policy*, 7(4):103–131, 2015.

G. Cassan and L. Vandewalle. Identities and public policies: Unexpected effects of political reservations for women in india. *World Development*, 143:105408, 2021.

L.J. Charleston, R.P. Adserias, N.M. Lang, and J.F.L. Jackson. Intersectionality and stem: The role of race and gender in the academic pursuits of african american women in stem. *Journal of Progressive Policy and Practice*, 2(3):273–293, 2014.

K.B. Coffman. Evidence on self-stereotyping and the contribution of ideas. *The Quarterly Journal of Economics*, 129(4):1625–1660, 2014.

O. Compte and A. Postlewaite. Confidence-enhanced performance. *American Economic Review*, 94(5):1536–1557, 2004.

O. Compte and A. Postlewaite. *Ignorance and Uncertainty.* Cambridge University Press, 2019.

K. Crenshaw. Mapping the margins: Intersectionality, identity politics, and violence agains women of color. *Stanford Law Review*, 43:1241–1299, 1991.

P.G. Davies, S.J. Spencer, D.M. Quinn, and R. Gerhardstein. Consuming images: how television commercials that elicit stereotype threat can restrain women academically and professionally. *Personal. Soc. Psychol. Bull.*, 28(12):1615–1628, 2002.

K. Eliaz and R. Spiegler. A model of competing narratives. *American Economic Review*, 110(12):3786–3816, 2020.

J. Elster. Rationality and economics. *The Economic Journal*, 106(438):1386–1397, 1996.

I. Esponda and M. Pouzo. Berk-nash equilibrium: A framework for modeling agents with misspecified models. *Econometrica*, 84(3):1093–1130, 2016.

C. Exley and J.B. Kessler. Motivated errors. Working paper 26595, NBER, 2019.

L. Festinger. A theory of social comparison processes. *Human Relations*, 7:117–140, 1954.

K. Fiedler and H. Bless. *Emotions and belief: How feelings influence thoughts*, chapter The formation of beliefs at the interface of affective and cognitive processes, pages 144–170. Cambridge University Press, 2000.

J.A. Flory, Leibbrandt A., and J.A. List. Do competitive work places deter female workers? a large-scale natural field experiment on gender differences in job-entry decisions. *The Review of Economic Studies*, 82(1):122–155, 2015.

O. Folke, L. Freidenvall, and J. Rickne. Gender quotas and ethnic minority representation: Swedish evidence from a longitudinal mixed methods study. *Politics & Gender*, 11(2):345–381, 2015.

D. Fudenberg and D.K. Levine. Self-confirming equilibrium. *Econometrica*, 61(3): 523–545, 1993.

N. Gupta. Intersectionality of gender and caste in academic performance: quantitative study of an elite indian engineering institute. *Gender, Technology and Development*, 23(2):165–186, 2019.

N. Guyon and E. Huillery. Biased aspirations and social inequality at school: Evidence from french teenagers. *The Economic Journal*, 131(634):745–796, 2021.

J. Henrich. *The Secret of Our Success.* Princeton University Press, 2016.

M.A. Hogg and P. Grieve. Social identity theory and the crisis of confidence in social psychology: A commentary, and some research on uncertainty reduction. *Asian Journal of Social Psychology*, 2:79–93, 1999.

M.M. Hughes. Intersectionality, quotas, and minority women's political representation worldwide. *American Political Science Review*, 105(3):604620, 2011.

R. Jia and T. Persson. Individual vs. social motives in identity choice: Theory and evidence from china. *ERN: Intrinsic Motivation (Topic)*, 2019.

V. Karekurve-Ramachandra and A. Lee. Do gender quotas hurt less privileged groups? evidence from india. *American Journal of Political Science*, 64(4):757–772, 2020.

Q. Lippmann and C. Senik. Math, girls and socialism. *Journal of Comparative Economics*, 46(3):874–888, 2018.

C.W. Liqui Lung. On the origin and persistence of identity-driven choice behavior. Working paper, Paris School of Economics, 2022.

B. Major and L. O'Brien. The social psychology of stigma. *Annuel Review of Psychology*, 56:393–421, 2005.

M. Mobius, M Niederle, P. Niehaus, and T. Rosenblat. Managing self-confidence. 2014.

F. Murden. *Mirror Thinking*. Bloomsbury Sigma, 2020.

J. Nagel. Constructing ethnicity: Creating and recreating ethnic identity and culture. *Social Problems*, 41(1):152–176, 1994.

T. O'Donoghue and M. Rabin. Doing it now or later. *American Economic Review*, 89 (1):103–124, 1999.

S. Oh. Does identity affect labor supply. *American Economic Review*, 113(8):2055–2083, 2023.

T. Perry, C. Steele, and A.G. Hilliard III. *Young, Gifted, and Black*. Beacon Press, 2003.

G. Pogrebna, S. Angelopoulos, and I. Motsi-Omoijiade. The impact of intersectional racial and gender biases on minority female leadership over two centuries. *Scientific Reports*, 14(111), 2024.

E. Pronin, C.M. Steel, and L. Ross. Identity bifurcation in response to stereotype threat: women and mathematics. *Journal of Experimental Social Psychology*, 40(2): 152–168, 2004.

N. Qian and E. Nix. The fluidity of race: Passing in the united states 1880-1940. Working paper, NBER, 2015.

E.A. Reynolds Losin, M. Iacoboni, A. Martin, and M. Dapretto. Own-gender imitation activates the brain's reward circuitry. *Social Cognitive and Affective Neuroscience*, 7(7):804–810, 2012.

Emily Rosenzweig. With eyes wide open: How and why awareness of the psychological immune system is compatible with its efficacy. *Perspectives on Psychological Science*, 11(2):222–238, 2016.

L. Ross and R.E. Nisbett. *The Person and the Situation.* McGraw-Hill, 1991.

S. Saccardo and M. Serra-Garcia. Enabling or limiting cognitive flexibility? evidence of demand for moral commitment. *American Economic Review*, 113(2):396–429, 2023.

J. Schwartzstein and A. Sunderarm. Using models to persuade. *American Economic Review*, 111(1):276–323, 2021.

M. Seligman. *Learned Optimism: How to change your mind and your life.* Random House USA Inc, 2006.

M. Shayo. A model of social identity with an application to political economy: nation, class, and redistribution. *American Political Science Review*, 103(02):147–174, May 2009.

J.L. Smith, C. Sansone, and P.H. White. The stereotyped task engagement process: the role of interest and achievement motivation. *Journal of Education and Psychology*, 99(1):99–114, 2007.

R. Spiegler. Bayesian networks and boundedly rational expectations. *The Quarterly Journal of Economics*, 131(3):1243–1290, 2016.

C.M. Steele. *Whistling Vivaldi: How stereotypes affect us and what we can do.* W.W. Norton and Company, Inc, 2010.

J. Steele, J.B. James, and R.C. Barnett. Learning in a mans world: examining the perceptions of undergraduate women in male-dominated academic areas. *Psychol. Women Q.*, 26(1):46–50, 2002.

N. Tan. Ethnic quotas and unintended effects on womens political representation in singapore. *International Political Science Review*, 35(1):27–40, 2014.

J. Van de Weele, E. Tripodi, and P. Schwardmann. Self-persuasion: Evidence from field experiments at international debating competitions. *American Economic Review*, 112(4):1118–1146, 2022.

N. Yuval-Davis. Intersectionality and feminist politics. *European Journal of Women's Studies*, 12(3):193–209, 2006.

N. Yuval-Davis. Situated intersectionality and social inequality. *Raisons Politiques*, 2 (58):91–100, 2015.

Appendix 1: Mathematical Appendix

PROPOSITION 1 (Individually Optimal Belief Formation): *The individually optimal strategies σ_i^* given an agent's type $\{\alpha_i, \Theta_i\}$ and a social context Π are the following:*

	$\alpha_i > \gamma$	$\alpha_i < \gamma$
$\overline{\eta}_\Theta > 1$	$\overline{\kappa}_\Theta$	R
$\underline{\eta}_\Theta < 1$	R	$\underline{\kappa}_\Theta$

Proof. Agents choose σ_i to maximize $V_i(\sigma_i|\alpha_i, \Theta_i, \Pi)$ over all possible realizations of $\hat{\alpha}_i$. Consider first agents with $\alpha_i > \gamma$. The welfare-maximizing choice for these agents is $a_i = C$. $V_i(\sigma_i|\alpha_i, \Theta_i, \Pi) > V_i(R|\alpha_i, \Theta_i, \Pi)$ for $\sigma_i \neq R$ if and only if $\Phi(\alpha_i, \Theta_i, \sigma_i, \Pi) > \Phi(\alpha_i, \Theta_i, \sigma_i, \Pi)$ for some $\sigma_i \in \{\theta^A, \theta^B\}$. Since $\Phi(\alpha_i, \Theta_i, \sigma_i, \Pi) = P(\hat{\alpha}_i > \gamma^{\sigma_i}|\alpha_i, \Theta_i, \Pi)$, this is the case when $\gamma^{\sigma_i} < \gamma$. This is true if and only if $\pi_{\theta_i^k} > \overline{\pi}_{\theta_i^k}$ for some $k \in \{A, B\}$. When multiple social identity cues satisfy this condition, agents maximize $V_i(\sigma_i|\alpha_i, \Theta_i, \Pi)$ by choosing σ_i to maximize $\gamma - \gamma^{\sigma_i}$. Hence, $\sigma_i^* = \overline{\kappa}_\Theta$. Otherwise, $\sigma_i^* = R$.

The proof is vice versa for agents with $\alpha_i < \gamma$. $V_i(\sigma_i|\alpha_i, \Theta_i, \Pi) > V_i(R|\alpha_i, \Theta_i, \Pi)$ if and only if $\Phi(\alpha_i, \Theta_i, \sigma_i, \Pi) < \Phi(\alpha_i, \Theta_i, \sigma_i, \Pi)$ for some $\sigma_i \in \{\theta^A, \theta^B\}$. This is the case if and only if $\gamma^{\sigma_i} > \gamma$, meaning that we need $\pi_{\theta_i^k} < \overline{\pi}_{\theta_i^k}$ for some $k \in \{A, B\}$. When multiple social identity cues satisfy this condition, agents maximize $V_i(\sigma_i|\alpha_i, \Theta_i, \Pi)$ by choosing σ_i to maximize $\gamma^{\sigma_i} - \gamma$. Hence, $\sigma_i^* = \underline{\kappa}_\Theta$. Otherwise, $\sigma_i^* = R$. □

PROPOSITION 2 (Potential to Improve Decision Making): *Asymmetry $\pi_{\theta^k} \neq \overline{\pi}_{\theta^k}$ along the lines of at least two observable traits θ^A and θ^B leads to inequalities in the potential to improve decision making across the different social types Θ. Specifically, agents with a mixed social type will have on average a higher expected pay-off $V_i(\sigma_i^*|\alpha_i, \Theta_i, \Pi)$ than agents with a one-sided social type.*

Proof. Agents with a *mixed* social type can use their social identity cues to bias $\hat{\alpha}_i$ both upwards and downwards. Agents with a *one-sided* social type can either bias $\hat{\alpha}_i$ upwards or downwards. Therefore, they can either correct of a Type I or Type II error, but not for both. Because $V_i(\sigma_i^*|\alpha_i, \Theta_i, \Pi) > V_i(R|\alpha_i, \Theta_i, \Pi)$ when $\sigma_i^* \neq R$, agents for whom it is optimal to not repress have a higher expected utility. For agents with a *mixed* social type this is the case both when $\alpha_i < \gamma$ and when $\alpha_i > \gamma$. For agents with a *one-sided* type, this condition only holds for either $\alpha_i < \gamma$ or $\alpha_i > \gamma$. For the other case, $V_i(\sigma_i^*|\alpha_i, \Theta_i, \Pi) = V_i(R|\alpha_i, \Theta_i, \Pi)$. Therefore, when we aggregate all agents along the lines of their social type Θ_i, conditional on σ^*, agents with a *mixed* social type have on average a higher expected payoff $V_i(\sigma_i^*|\alpha_i, \Theta_i, \Pi)$ than agents with a *one-sided* social type. □

COROLLARY 1: *Let $\pi_{\theta^k} > \overline{\pi}_{\theta^k}$ for all $k \in \{A, B\}$. We have a **type-specific population effect**, such that $\Phi(\alpha_i, \tilde{t}_{\theta^k}, \sigma_i^*, \Pi) > \Phi(\alpha_i, t \in T_{mixed}, \sigma_i^*, \Pi) > \Phi(\alpha_i, \tilde{t}_{\theta^{k'}}, \sigma_i^*, \Pi)$, and a **type-specific selection effect**, such that $E(\alpha_i|a_i = C, \tilde{t}_{\theta^k}) < E(\alpha_i|a_i = C, t \in T_{mixed}) < E(\alpha_i|a_i = C, \tilde{t}_{\theta^{k'}})$.*

Proof. If $\pi_{\theta^k} > \overline{\pi}_{\theta^k}$ for all $k \in \{A, B\}$, then for agents with social type $\tilde{t}_{\theta^k}$, we have $\overline{\eta}_{\tilde{t}_{\theta k}} > 1$ and $\underline{\eta}_{\tilde{t}_{\theta k}} > 1$. Hence, when $\alpha_i > \gamma$, $\sigma_i^* = \overline{\kappa}_{\tilde{t}_{\theta k}}$, and when $\alpha_i < \gamma$, we have $\sigma_i^* = R$. For all $\Theta \in T_{mixed}$, we have $\overline{\eta}_\Theta > 1$ and $\underline{\eta}_\Theta < 1$. Therefore, $\sigma^* = \overline{\kappa}_\Theta$ when $\alpha_i > \gamma$ and $\sigma_i^* = \underline{\kappa}_\Theta$ when $\alpha_i < \gamma$. According to Definition 1, when $\sigma^* \neq R$, then for agents with $\alpha_i > \gamma$, we have $\Phi(\alpha_i, \Theta_i, \sigma_i^*, \Pi) > \Phi(\alpha_i, \Theta_i, R, \Pi)$. Similarly, for agents with $\alpha_i > \gamma$, we have $\Phi(\alpha_i, \Theta_i, \sigma_i^*, \Pi) < \Phi(\alpha_i, \Theta_i, R, \Pi)$. Now, it follows that $\Phi(\alpha_i < \gamma, \tilde{t}_{\theta^k}, \sigma_i^*, \Pi) > \Phi(\alpha_i < \gamma, t, \sigma_i^*, \Pi)$ for all $t \neq \tilde{t}_{\theta^k}$, while $\Phi(\alpha_i > \gamma, \tilde{t}_{\theta^{k'}}, \sigma_i^*, \Pi) > \Phi(\alpha_i < \gamma, t, \sigma_i^*, \Pi)$ for all $t \neq \tilde{t}_{\theta^{k'}}$. Hence, aggregating agents over all ability types $\alpha_i \in [0, 1]$ along the lines of social types t results in $\Phi(\alpha_i, \tilde{t}_{\theta^k}, \sigma_i^*, \Pi) > \Phi(\alpha_i, t \in T_{mixed}, \sigma_i^*, \Pi) > \Phi(\alpha_i, \tilde{t}_{\theta^{k'}}, \sigma_i^*, \Pi)$.

If $\Phi(\alpha_i, \tilde{t}_{\theta^k}, \sigma_i^*, \Pi) > \Phi(\alpha_i, t \in T_{mixed}, \sigma_i^*, \Pi) > \Phi(\alpha_i, \tilde{t}_{\theta^{k'}}, \sigma_i^*, \Pi)$, then on average $\gamma_{\sigma_i^*}$ for agents with type $\tilde{t}_{\theta^k}$ is lower than the average $\gamma_{\sigma_i^*}$ for agents with $\Theta_i \in T_{mixed}$. And the latter thresholds are on average lower than the thresholds for agents with type $\tilde{t}_{\theta^{k'}}$. Hence, agents with type $\tilde{t}_{\theta^{k'}}$ choose $a_i = C$ for on average the highest realizations of $\hat{\alpha}_i$. Because, according to Assumption 1, for each agent $E(\hat{\alpha}_i) = \alpha_i$, these agents will on average have higher ability types α_i. It then follows that $E(\alpha_i | a_i = C, \tilde{t}_{\theta^k}) < E(\alpha_i | a_i = C, t \in T_{mixed}) < E(\alpha_i | a_i = C, \tilde{t}_{\theta^{k'}})$. □

PROPOSITION 3: *If a* ***Symmetric Equilibrium*** *is unstable in both dimensions* $k \in \{A, B\}$*, then it co-exists with a stable* ***Asymmetric Equilibrium of Degree 2****. Assume WLOG that in any* ***Asymmetric Equilibrium of Degree 2*** $\pi_{\theta^k} > \overline{\pi}_{\theta^k}$ *for* $k \in \{A, B\}$*. Then, with* $S_\Theta = p_\Theta \int \alpha \Phi(\alpha, \Theta, \sigma^*, \Pi) d\alpha$*, the order on* S_Θ *must be such that,*

$$S_{\tilde{t}_{\theta^k}} > S_{T_{mixed}} > S_{\tilde{t}_{\theta^{k'}}}$$

Proof. We infer a *Symmetric Equilibrium* always exists. When $\Pi : \pi_{\theta^k} = \overline{\pi}_{\theta^k}$ for all $k \in \{A, B\}$, then, all strategies $\sigma \in \{\theta^A, \theta^B, R\}$ are equivalent. Since α_i and Θ_i are independently distributed over the population, $\tilde{\pi}_{\theta^k}(\Pi, \sigma) = \overline{\pi}_{\theta^k}$ for all $k \in \{A, B\}$.

Now consider a perturbation of a *Symmetric Equilibrium* such that $\pi^\delta_{\theta^k} = \overline{\pi}_{\theta^k} + \delta$, while $\pi^\delta_{\theta^{k'}} = \overline{\pi}_{\theta^{k'}} - \delta$ for all $k \in \{A, B\}$. Let $\mathcal{S}^{sym}$ be the set of all social contexts Π in which the induced social identity cues $\tilde{\pi}_{\theta^k}(\Pi, \sigma)$ are symmetric for $k \in \{A, B\}$. Similarly, we can now define the set $\mathcal{S}^{sym}_\delta$ that contains all Π that are induced by a perturbation δ such that $\pi^\delta_{\theta^k} = \overline{\pi}_{\theta^k} + \delta$, while $\pi^\delta_{\theta^{k'}} = \overline{\pi}_{\theta^{k'}} - \delta$ for all $k \in \{A, B\}$. It follows that $\mathcal{S}^{sym}_\delta \subset \mathcal{S}^{sym}$ and we have a non-empty, compact and convex set $\mathcal{S}^{sym}_\delta$.

Secondly, we know $\tilde{\pi}_{\theta^k}(\sigma^*, \Pi)$ is some linear transformation of,

$$\int_{\alpha > \gamma} \alpha P\left(\hat{\alpha} > \frac{\gamma}{\eta(\pi_{\theta^k}, \overline{\pi}_{\theta^k})}\right) f(\hat{\alpha}|\alpha) d\hat{\alpha} d\alpha + \int_{\alpha < \gamma} \alpha P(\hat{\alpha} > \gamma) f(\hat{\alpha}|\alpha) d\hat{\alpha} d\alpha \qquad (6)$$

When $\eta(\pi, \overline{\pi})$ then $\tilde{\pi}_{\theta^k}(\sigma^*, \pi)$ is continuous in π. Furthermore, $\tilde{\pi}_{\theta^k}(\sigma^*, \pi)$ is increasing in π. Then, when there exist $\hat{\pi} > \overline{\pi}$ and θ^k such that $\tilde{\pi}_{\theta^k}(\sigma^*, \hat{\pi}) > \hat{\pi}$, then $\forall \pi > \hat{\pi}$, we have $\tilde{\pi}_{\theta^k}(\sigma^*, \pi) > \hat{\pi}$. This condition holds when a *Symmetric Equilibrium* is unstable according to Definition 5. If this is the case, we now have a non-empty, compact and convex set $\mathcal{S}_\delta^{sym}$, and a continuous function $\tilde{\Pi}(\sigma, \cdot) : \mathcal{S}_\delta^{sym} \to \mathcal{S}_\delta^{sym}$. Therefore, following Brouwer's fixed point theorem, there exists a fixed point $\Pi^* \in \mathcal{S}_\delta^{sym}$ such that, $\Pi^* = \tilde{\Pi}(\sigma, \mathcal{S}_\delta^{sym})$. According to Definition 4 and the definition of $\mathcal{S}_\delta^{sym}$, such a fixed point Π^* is an *Asymmetric Equilibrium of Degree 2.*

Finally, because for any $\Pi^* \in \mathcal{S}_\delta^{sym}$, $\pi_{\theta^k}^\delta > \overline{\pi}_{\theta^k}$, while $\pi_{\theta^{k\prime}}^\delta < \overline{\pi}_{\theta^{k\prime}}$ for all $k \in \{A, B\}$, from Corollary 1 it follows we necessarily have that in any such fixed point Π^*,

$$S_{\tilde{t}_{\theta k}} > S_{T_{mixed}} > S_{\tilde{t}_{\theta k\prime}}$$

□

COROLLARY 2: *Take two response functions $\hat{\eta}$ and η, such that $\hat{\eta}(\pi, \overline{\pi}) > \eta(\pi, \overline{\pi})$ for all $\pi > \overline{\pi}$. Assume WLOG that an 'Asymmetric Equilibrium' exists in which $\pi > \overline{\pi}$. Let π_η^* be the equilibrium value of π given a response function η. Then, $\pi_{\hat{\eta}}^* > \pi_\eta^*$.*

Proof. Let $\eta(\pi, \overline{\pi})$ be a response function such that, given γ, the condition of Lemma 1 holds. Then, an Asymmetric Equilibrium also exists for any response function $\hat{\eta}(\pi, \overline{\pi})$, such that $\hat{\eta}(\pi, \overline{\pi}) > \eta(\pi, \overline{\pi})$ for all $\pi > \overline{\pi}$. Let $\tilde{\pi}_{\eta,\theta}(\sigma, \pi)$ be the induced value of π for a response function η. Then, if $\hat{\eta}(\pi, \overline{\pi}) > \eta(\pi, \overline{\pi})$ for all $\pi > \overline{\pi}$, we have $\tilde{\pi}_{\hat{\eta},\theta}(\sigma, \pi) > \tilde{\pi}_{\eta,\theta}(\sigma, \pi) \;\; \forall \pi > \overline{\pi}$. Consequently, let π_η^* be the equilibrium value of π that arises in an Asymmetric Equilibrium for a response function η. Then, $\pi^{(1)} \equiv \tilde{\pi}_{\hat{\eta},\theta}(\sigma, \pi_\eta^*) > \tilde{\pi}_{\eta,\theta}(\sigma, \pi_\eta^*) = \pi_\eta^*$, which implies that $\pi^{(2)} \equiv \tilde{\pi}_{\hat{\eta},\theta}(\sigma, \pi^{(1)}) > \tilde{\pi}_{\hat{\eta},\theta}(\sigma, \pi_\eta^*) \equiv \pi^{(1)}$ and $\pi^{(3)} \equiv \tilde{\pi}_{\hat{\eta},x}(\sigma, \pi^{(2)}) > \tilde{\pi}_{\hat{\eta},1}(\sigma, \pi^{(1)}) \equiv \pi^{(2)}$. This sequence converges to $\pi_{\hat{\eta}}^* = \tilde{\pi}_{\hat{\eta},\sigma}(\sigma, \pi_{\hat{\eta}}^*)$ and is everywhere above π_η^* and below the upper bound π^u on π. This shows that, for any response function $\hat{\eta}(\pi, \overline{\pi})$ such that $\hat{\eta}(\pi, \overline{\pi}) > \eta(\pi, \overline{\pi})$ for all $\pi > \overline{\pi}$, in equilibrium

$$\pi_{\hat{\eta}}^* > \pi_\eta^* \tag{7}$$

□

COROLLARY 3: *The individually optimal strategies σ_i^* given an agent's type $\{\alpha_i, \Theta_i\}$ with $\sigma_i \in \{\theta^A, \theta^B, \Theta, R\}$ and a social context Π are the following:*

	$\alpha_i > \gamma$	$\alpha_i < \gamma$
$\overline{\overline{\eta}}_\Theta > 1$	$\overline{\overline{\kappa}}_\Theta$	R
$\underline{\underline{\eta}}_\Theta < 1$	R	$\underline{\underline{\kappa}}_\Theta$

Proof. The proof of this corollory is analogous to the proof of Proposition 1. Agents choose $\sigma_i \in \{\theta^A, \theta^B, \Theta, R\}$ to maximize $V(\sigma_i|\alpha_i, \Theta_i, \Pi)$. Hence they will choose $\sigma_i = \Theta$ if and only if $\Theta = \text{argmax}_{\sigma_i \in \{\theta^A, \theta^B, \Theta, R\}} V(\sigma_i|\alpha_i, \Theta_i, \Pi)$. When $\alpha_i > \gamma$, maximizing $V(\sigma_i|\alpha_i, \Theta_i, \Pi)$ is equivalent to maximizing $\Phi(\alpha_i, \Theta_i, \sigma_i, \Pi)$. We only have $\Phi(\alpha_i, \Theta_i, \Theta, \Pi) = \max_{\sigma_i \in \{\theta^A, \theta^B, \Theta, R\}} \Phi(\alpha_i, \Theta_i, \sigma_i, \Pi)$ when $\pi_\Theta > \overline{\pi}_\Theta$ and $\Theta = \overline{\overline{\kappa}}_t$. Vice versa, when $\alpha_i < \gamma$. □

Appendix 2: Strategy Restrictions

The Three-Strategy Model - Agents can use both *one-dimensional* and *two-dimensional* social identity cues. When gender is stigmatized, the strategy set becomes $\sigma_i \in \{Gender, \Theta, R\}$. There are two different settings. In the first setting, gender is *dominant*, and both female students with an Asian and Western last name are relatively underrepresented among those qualified for the international final. In this case, both *talented* female students with an Asian last name and *not talented* male students with a Western last name cannot use $\sigma_i = \Theta$ to improve decision making. Consequently, the *Three-Strategy* model has similar implications as the *Two-Strategy* model, where *mixed* social types lose their ability to decrease the likelihood of making one type of mistake. When agents with *one-sided* social types find it optimal to use π_Θ instead of π_{Gender}, the optimal strategy of agents with *one-sided* social types induces a larger bias than the optimal strategy of agents with *mixed* social types. This induces differences in choice behavior across students with different last names. Unlike in the *Two-Strategy Model*, the availability of *two-dimensional* cues can therefore induce population and selection effects in dimensions of the social type that are not stigmatized.

In the second setting, the *Name*-dimension is *dominant*, and both female and male agents with a Western last name are relatively underrepresented among those successful. Here, students with a *mixed* social type maintain their ability to potentially correct for both types of mistakes. In this setting, the availability of *two-dimensional* cues can therefore enable agents with *mixed* social types to escape the negative effects of the stigmatization of *Gender*. There will be a difference in the potential to improve decision making across agents with *mixed* social types. To determine what happens in equilibrium, we can use Corollary 3.

The Asymmetric Model - Assume WLOG that being female is stigmatized. This means that for female students the strategy set is restricted to $\sigma_i \in \{\theta^A, \Theta, R\}$. Male students, on the other hand, can use the complete strategy set $\sigma_i \in \{\theta^A, \theta^B, \Theta, R\}$. First, consider the setting in which the dimension of last name is *dominant*. In this setting, female students with an Asian last name can potentially escape the negative effects of stigmatization by using their *two-dimensional* cue π_{FA}. If it is optimal to choose $\sigma_i = k$ for all agents with a *mixed* social type, then stigmatization only negatively affects *talented* female students with an Asian last name. As a result, the potential of male students and students with an Asian last name to improve decision making slightly increases, while it slightly decreases for female students and students with a Western last name. Hence, the population and selection effects in the dimension of gender will be reinforced, while the strength of the population and selection effects in the dimension of last names decreases. When gender is *dominant*, talented female students with an Asian last name are no longer able to potentially correct for Type-II error. This decreases the participation of both female students and Asian students at the same time, increasing the population and selection effects in the dimension of gender, while decreasing them in the dimension of last names.