UNIVERSITY OF CAMBRIDGE
Faculty of Economics

JANEWAY INSTITUTE CAMBRIDGE

CAMBRIDGE WORKING PAPERS IN ECONOMICS
JANEWAY INSTITUTE WORKING PAPERS

Intersectionality in Individual Choice Behavior: Pitfalls and Opportunities

Caroline
Liqui-Lung
University
of Cambridge

Abstract

I show how intersectionality, interconnections of social organizations that create interdependent systems of disadvantage, plays a role in individual choice behaviour when people use outcomes of others like them to cope with sources of noise in decision making they cannot control for. I analyze how the different dimensions of a social type interact in belief formation and choice behaviour at the individual and aggregate level, and show how an intersectional lens sheds light on inequalities and patterns in aggregate choice behaviour that are not visible with a one-dimensional lens. I furthermore discuss the effects of strategy restrictions imposed by stigmatization, stereotypes or norms, and the ability of agents to self-identify. Finally, I illustrate how these insights could help explain the pitfalls we encounter in the evaluation of one-dimensional policy measures targeting the underrepresentation of social groups, and guide us in developing potentially more effective multidimensional approaches.

Reference Details

2379	Cambridge Working Papers in Economics
2321	Janeway Institute Working Paper Series
Published	06 December 2023
Revised	17 January 2024
JEL Code	D81, D91, Z13
Websites	www.econ.cam.ac.uk/cwpe
	www.janeway.econ.cam.ac.uk/working-papers

Intersectionality in Individual Choice Behaviour: The Pitfalls and Opportunities

Caroline Liqui-Lung *

Version: January 17, 2024

I show how intersectionality, interconnections of social organizations that create interdependent systems of disadvantage, plays a role in individual choice behaviour when people use outcomes of others like them to cope with sources of noise in decision making they cannot control for. I analyze how the different dimensions of a social type interact in belief formation and choice behaviour at the individual and aggregate level, and show how an intersectional lens sheds light on inequalities and patterns in aggregate choice behaviour that are not visible with a one-dimensional lens. I furthermore discuss the effects of strategy restrictions imposed by stigmatization, stereotypes or norms, and the ability of agents to self-identify. Finally, I illustrate how these insights could help explain the pitfalls we encounter in the evaluation of one-dimensional policy measures targeting the underrepresentation of social groups, and guide us in developing potentially more effective multidimensional approaches.

(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, Willemien Kets, Jeanne Hagenbach, Glenn Ellison and the participants of the EEA Barcelona conference and seminars at the University of Cambridge and the Paris School of Economics for their useful comments and suggestions.

1 Introduction

Fostering social diversity in professional, political and educational environments is essential for advancing social mobility, productivity and economic development. However, recent literature challenges the effectiveness of traditional one-dimensional policy measures, such as those centered on gender alone, showing such measures may have limited impact (Breda et al., 2023) or can have unintended spillover effects on the representation of other social groups, defined by for example caste (Cassan and Vandewalle, 2021). Further evidence indeed suggests that the representation of different social groups interact. Women's career trajectories flatten mid-career compared to men's, where this downward trend is strongest for women belonging to ethnic minorities[1]. Similarly, Gupta (2019) shows how gender has a different effect on academic performance for men and women belonging to different castes. These interconnections of social organizations and the idea that they create interdependent systems of disadvantage are commonly referred to with the term "intersectionality" (Crenshaw, 1991), and the resulting effects go beyond simply adding the separate effects different dimensions of one's social identity induce (Yuval-Davis, 2015).

This paper explores how intersectionality can both undermine one-dimensional policies and can be harnessed to develop more effective multidimensional measures. Although previous studies focus on intersectionality as a result of interactions between agents in the context of discrimination and access to public goods[2], another important observation challenging diversity is that a priori identical individuals from different social groups make different occupational and educational choices. This study focuses therefore on individual decision making, and introduces a novel source of intersectionality, highlighting the role of boundedly rational agents using information about the outcomes of others to navigate sources of noise that are beyond their control.

The paper builds on Liqui Lung (2022), that shows how individuals that are not able to correct for all noise in decision making as a Bayesian would, can improve decision making by using statistics about the prevalence of their social group among those successful to bias their decision making in a direction contingent on their social

[1] See the 'Leaders and Daughters Global Survey 2019' for more details

[2] See e.g. Crenshaw (1991), Yuval-Davis (2006), Rebughini (2021)

type. This *optimal self-screening* can nevertheless induce persistent identity-driven choices. Hence, when agents' cognitive abilities are restricted, informationally irrelevant social context can turn social inequalities into a self-fulfilling prophecy. In this paper, I extend the model to allow for multi-dimensional social types, allowing us to simultaneously consider for example an agent's gender, ethnicity or social class. I analyze how the different dimensions of a social type interact in choice behaviour at both the individual and the aggregate level. I furthermore discuss the effects of strategy restrictions imposed by stigmatization, stereotypes or norms, and the ability of agents to self-identify. I use the insights to shed light on both the pitfalls and opportunities intersectionality can create for the development of policy to foster social diversity.

The mechanism builds on two ideas. First, research in social psychology underscores how individuals' optimism or pessimism about their own chances of success in tasks are influenced by the outcomes of others.[3]. Secondly, although we may generally have an accurate perception of our abilities, exogenous factors, such as emotions or recent feedback, can induce noise in decision making, making us momentarily over- or under confident (see e.g. Fiedler and Bless (2000) and Elster (1996)). This noise can drive us to making sub-optimal choices in tasks related to these abilities.

Consider students choosing whether to enter a math competition. The noise in their perception of their chances of success can induce two mistakes. First, they may be momentarily too optimistic and enter the competition, while this is not their welfare-maximizing choice (Type I error). Second, they may be momentarily too pessimistic and not enter the competition, while this is their welfare-maximizing choice (Type II error). At the same time, assume students observe data about those successful in the competition in the previous year, with their gender and whether they belong to another underrepresented minority (URM). In the model, boundedly rational agents can use this information, which I will refer to as *social identity cues*, as an instrument to potentially limit the negative effects of noise on decision making. Specifically, when forming a belief about their probability of success in a task, they can either naively follow their own noisy perception of this probability, or use statistics about

[3]See the work of Seligman (2006) who shows people can take the successes and failures of others like them as evidence they will fail or succeed as well, and Murden (2020) who shows how our behaviour and beliefs are influenced by the choices and outcomes of others through 'imitation'.

the prevalence of their social group among those successful to bias this noisy perception in a direction contingent on their social type. For example, assume male and non-URM students were relatively overrepresented among those successful in the previous year. Hence, a male URM student can use the cue related to gender to make himself more optimistic, while he can use the cue related to belonging to an URM to make himself more pessimistic about his own chances of success in the competition.

I show how agents endogenously determine on which dimension of their social type they focus as a function of their fitness with respect to a task and whether their different social groups are relatively under- or overrepresented among the successful individuals. For example, when it is not optimal for a male URM student to enter the math competition, he can minimize the ex-ante likelihood of making a type I error by biasing his noisy perception downward with the cue related to the underrepresentation of URM students. Hence, he learns to ignore the *social identity cue* related to gender. On the other hand, if this student is very good at math, he can minimize the ex-ante likelihood of making a type II error by biasing his noisy perception upward using the cue related to gender, and learns to ignore the cue related to belonging to an URM.

The key insight of the paper is that the advantages induced by a multi-dimensional social type transcend the mere representation of the respective social groups among the successful individuals, and instead lie in the ability of a social type to bias a noisy perception optimally. Specifically, there are two categories of social types. There are *mixed* social types, e.g. non-URM female and URM male students, who belong to the socially more successful group according to one dimension of their social type, while they belong to the socially less successful group according to another dimension. Secondly, there are *one-sided* social types, e.g. URM female students and non-URM male students, who belong to either the socially more or less successful group according to all dimensions of their social type. URM female students are unable to bias their noisy perception upwards. They are therefore more likely to make a Type I error. Non-URM male students, on the other hand, cannot bias their noisy perception downwards, and are therefore more likely to make a Type II error. URM male and non-URM female students, on the other hand, can bias their perception both upwards and downwards. They can therefore decrease the likelihood of making both types of error. Intersectionality in individual choice behaviour induces therefore on average a higher expected utility for agents with *mixed* social types than for agents with *one-sided* social types.

This mechanism induces patterns in aggregate choice behaviour that are in line with what we observe in real settings.[4] As URM female students are most likely to make a Type I error, they have the smallest propensity to enter the competition. On the other hand, non-URM male students are most likely to make a Type II error, and have therefore the largest propensity to enter the competition. At the same time, conditional on entering the competition, URM female students will have a higher success rate than non-URM male students. I consequently show how this results in Non-URM male students being most overrepresented among the successful students in the next year, while URM female students will be most underrepresented. The representation of non-URM female students and URM male students will be in between these two groups. An equilibrium analysis shows moreover how these differences in choice behaviour and representation can be persistent. Analyzing data using an *intersectional* lens enables us therefore to observe inequalities that are not visible when we only use a *one-dimensional* lens. Finally, asymmetric choices in a particular dimension of the social type, for example gender, are not equally driven by all male and female agents, but predominantly by the behaviour of agents with a *one-sided* social type.

To derive further policy-relevant insights, I analyze the effects of strategy restrictions. When agents are unable to ignore cues related to a particular dimension of their social type, because this dimension is stigmatized or there exist stereotypes or norms that make this dimension of identity salient, we observe larger asymmetries in choice behaviour along the lines of this dimension of the social type than along the lines of the 'non-stigmatized' dimension. Such strategy restrictions particularly affect the welfare of agents with a *mixed* social type, but under certain circumstances these negative effects can be escaped when agents have access to *two-dimensional* cues, meaning cues regarding for example URM female students in particular. On the other hand, the more control agents have to determine their social type, the more this will decrease the size of the population with a *one-sided* social type, and hence the asymmetries in

[4] An example are the results over the years in US (https://mathprize.atfoundation.org/experience/past-events) and international math competitions (https://www.egmo.org/egmos/egmo12/scoreboard/), focusing on gender and being of Asian descent.

choice behaviour. This may be an intuition for why we tend to see persistent under-representation of groups predominantly along the lines of dimensions of social identity that are difficult or costly to manipulate.

To illustrate how the insights this model provides can be relevant for policy makers, I discuss as example the use of quota to increase the representation of female students in STEM careers. Specifically, within the particular framework presented in this paper, I analyze how a one-time quota can affect the decision to apply to a STEM career in the next generation. I show how an intersectional view and knowledge of the strategy restrictions prospective students are subject to are crucial for the development of this policy. Without it, we may be providing suboptimal statistical cues or data, the policy may not affect the students we target or it can transfer asymmetries along the lines of one dimension of social identity to another. Finally, I discuss how a combination of multidimensional quota and informational policy could help overcome the pitfalls of one-dimensional policy measures and improve its effectiveness.

The idea that people use the multidimensionality of their social identity to enhance utility also appears in other decision-making settings. In Benabou and Tirole (2011), identity investments serve as self signals when memory is imperfect. This helps agents to enhance expected utility through functional or affective benefits. Atkin et al. (2021) shows how ethnic and religious identities in India are determined by group status, group salience and the market cost of following a group's prescribed behaviours. Shayo (2009) discusses how group identification is driven by group status and perceived similarity along the different dimensions of social identity. This paper proposes a novel mechanism, where agents use the different dimensions of social identity to limit the adverse effects of noise that is beyond their control to improve decision making on average. Furthermore, I show how an agent's self-image or mental model (Hoff and Stiglitz, 2016) in a particular social context is determined endogenously through its instrumental value in decision making, and why people focus on different dimensions of their social identity in different contexts. Finally, Carvalho and Pradelski (2022) presents a related equilibrium model with multidimensional social identities. In their model, the representation of an agent's social group affects choice behaviour through a direct identity-based payoff in the utility function. Furthermore, in their model, the degree to which agents focus on a certain dimension of their social identity is exogenous. In this model, the representation of an agent's social group has no direct effects on utility. Instead, both the use of social identity in belief formation and the degree to which agents focus on a certain dimension of social identity are determined endogenously, and are driven by a fitness criterion to improve decision making.

The rest of the paper is structured as follows. Section 2 presents the model. Section 3 presents the main results. Section 4 discusses strategy restrictions and self-identification. Section 5 presents the policy example. Finally, section 6 concludes. All formal proofs can be found in the Appendix.

2 The Model

2.1 The Environment

The baseline model will be the same as in Liqui Lung (2022). I consider a society with $i = 1, ..., N$ agents, with N arbitrarily large. Each agent i chooses an action $a_i \in \{C, NC\}$, where C and NC represent classes of tasks of a *Competence-Driven* and a *Non-Competence-Driven* type. The outcome of a_i can be either *'success'* or *'failure'* and is represented by $Y_i \in \{1, 0\}$. The probability of success for a *Competence-Driven* task depends on an agent's individual characteristics. This probability is represented by the continuous variable $\alpha \in [0, 1]$, and is distributed over the population following a distribution f_α. For each agent i, the probability of a successful outcome $Y_i = 1$ conditional on choosing the *Competence-Driven* task is fixed and given by,

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

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,

$$p(Y_i = 1|a_i = NC) = \gamma \tag{2}$$

Noisy Perceptions - The main assumption in the model is that agents only have a noisy perception $\hat{\alpha}_i$ regarding their own probability of success α_i. I pose that $\hat{\alpha}_i$ stems from a distribution g_{α_i} with $E(\hat{\alpha}_i) = \alpha_i$.[5]

Social Context - The novelty in this paper is that each agent is described by multiple *observable characteristics*. These observable characteristics can represent for example the agent's gender, ethnicity, social class or age, and are public information. To simplify to exposition of the model, I consider agents with two-dimensional social types $\Theta_i = (\theta_i^k)_{k\in\{A,B\}}$, where each θ_i^k is a binary characteristic with realizations $x \in \{0, 1\}$. For simplicity, I assume these characteristics are independently distributed over the population. Let $t \in T = \{11, 10, 01, 00\}$ be a possible realization of the social type Θ_i. I let p_x^k be the fraction of the population with observable characteristic $\theta^k = x$, and p_t is the fraction of the population with social type $\Theta = t$. Each agent i is fully described by (α_i, Θ_i). To isolate the mechanism through which social identity affects choice behaviour in this model, I assume the probabilities α and the social types Θ are independently distributed over the population.[6]

Agents have access to public data that consists of the outcome variables and the observable characteristics of other agents that have already made the choice. For the exposition of the model, I will focus on one particular statistic. In Liqui Lung (2022), I discuss how different data and different structures on information affect behaviour. Let $\mathcal{N}_{C,x}^k = \{i \in N, \theta_i^k = x, a_i = C\}$ be the set of agents with $\theta_i^k = x$ that have chosen the *Competence-Driven* task. Similarly, let $\mathcal{N}_{C,t} = \{i \in N, \Theta_i = t, a_i = C\}$ be the set of agents with social type $\Theta_i = t$ that have chosen the *Competence-Driven* task. Finally, 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 $\mathcal{N}_{C,x}^k, \mathcal{N}_{C,t} \subset \mathcal{N}_C$. Society then provides the

[5]As in Liqui Lung (2022), this noise should be interpreted as the effects of momentary emotions or distractions that agents do not have to tools to correct for. If agents could switch this noise off, they would behave as Bayesians. The assumption that the perception is unbiased only serves to show that a systematic bias is not what drives the results in this model.

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

statistics,

$$\pi_x^k = \frac{\sum_{i\in\mathcal{N}_{C,x}^k} Y_i}{\sum_{i\in\mathcal{N}_C} Y_i} \qquad\qquad \pi_t = \frac{\sum_{i\in\mathcal{N}_{C,t}} Y_i}{\sum_{i\in\mathcal{N}_C} Y_i}$$

for all $k \in \{A, B\}$, $x \in \{0, 1\}$ and $t \in \{11, 10, 01, 00\}$. These are the fractions of successful individuals with characteristic $\theta_i^k = x$ or social type $\Theta_i = t$ among all successful individuals that have chosen the *Competence-Driven* task. I call these fractions the '*social identity cues*' for agents with social type $\Theta_i = (\theta_i^A, \theta_i^B)$. Furthermore, I will refer to π_x^k as '*one-dimensional social identity cues*', and to π_t as '*two-dimensional social identity cues*'. Finally, I will refer to '*social context*' as the vector $\Pi = (\pi_t)_{t\in T}$.

Subjective Belief Formation - Because α_i, Θ_i are independently distributed over the population, social context contains no relevant information to learn about α_i. I assume nevertheless that agents have a natural 'urge' to look at others like them when they are not sure what to do, and have the option to either *Repress* or *Not Repress* this urge. Like in Liqui Lung (2022), I introduce a family of belief formation processes with which agents form a subjective belief $\hat{p}_i$ about their probability of success in a *Competence-Driven* task α_i, and I assume agents have some discretion in finding out which belief formation process suits them best. Going from one- to two-dimensional social types implies that this family of belief formation processes becomes larger. Specifically, when agents do not *Repress* the urge to look at others, they decide whether they consider others like them to be defined as agents with their same observable characteristic θ_i^A, their observable characteristic θ_i^B or as agents with their entire social type Θ_i.

To simplify the exposition of the model, I first consider on agents that only focus on *one-dimensional* social identity cues[7]. Hence, agents choose a strategy $\sigma_i \in \{A, B, R\}$ that results in a belief about α_i equal to $\hat{p}_i^{\sigma_i} \in \{\hat{p}_i^A, \hat{p}_i^B, \hat{p}_i^R\}$. Specifically, let η be a '*response function*' that is non-decreasing, such that,

$$\eta(\pi, p) = \begin{cases} > 1 & \text{if } \pi > p \\ 1 & \text{if } \pi = p \\ < 1 & \text{if } \pi < p \end{cases} \tag{3}$$

[7]Section 3.3 discusses the effects of agents also using *two-dimensional* social identity cues.

and let $\eta_{k,x} \equiv \eta(\pi_x^k, p_x^k)$. Then,

$$\hat{p}_i^{\sigma_i} = \begin{cases} \hat{\alpha}_i & \text{if } \sigma_i = R \\ \eta_{k,x}\hat{\alpha}_i & \text{if } \sigma_i \in \{A, B\} \end{cases} \tag{4}$$

As in Liqui Lung (2022) agents bias their noisy perception in a direction contingent on their social type. Here, they can nevertheless decide which dimension of their social type determines this direction. When the observable characteristic agents focus on implies belonging to the socially more successful subgroup, choosing $\sigma_i = k$ leads to an optimistic interpretation of $\hat{\alpha}$, while this leads to a pessimistic interpretation when the characteristic implies belonging to the socially less successful subgroup.

Subjective Utility Maximization - Agents derive utility from being successful and the utility function can be represented by $u_i = Y_i$. Each agent chooses her action a_i to maximize $E(u_i)$ given her subjective belief $\hat{p}_i^{\sigma}$, and chooses the *Competence-Driven* task if and only if $\hat{p}_i^{\sigma} > \gamma$. As in Liqui Lung (2022), the model can also be directly specified in terms of a choice set. Formally, subjective expected utility maximization implies that the agent is effectively comparing thresholds, such that agent i chooses $a = C$ if and only if $\hat{\alpha}_i > \gamma_i$, where

$$\gamma_i = \begin{cases} \gamma & \text{when } \sigma_i = R \\ \frac{\gamma}{\eta_{k,x}} & \text{when } \sigma_i \in \{A, B\} \end{cases} \tag{5}$$

The use of social identity cues implies agents inflate or deflate the threshold for $\hat{\alpha}$ above which they think they are 'good enough' to undertake the *Competence-Driven* task. This choice set in terms of γ_i is different for agents with different social types. This is the key driver of the results.

2.2 The Solution Concept

The effect of social context on belief formation affects the choices of task. This leads to outcomes that induce social identity cues, that in turn affect the way agents form subjective beliefs. To tractably capture the fixed points in this dynamic process, I use the same static solution concept as in Liqui Lung (2022), in which, given a social context, agents choose their strategy σ according to an individual optimality criterion I introduce below. I then define a population equilibrium as a fixed point in social context that is induced by the individually optimal strategy choices.

Individual Optimality - Let $\Phi_{\alpha,t,\sigma_i,\Pi} = P(a = C|\alpha, t, \sigma_i, \Pi)$ be the induced probability that an agent with α and social type $\Theta_i = t$ playing strategy σ_i given a social context Π chooses the *Competence-Driven* task. Then,

$$\Phi_{\alpha,t,\sigma_i,\Pi} = P(\hat{p}_i^{\sigma} > \gamma|\alpha) \tag{6}$$

This probability Φ follows from the distribution $g_{\alpha_i}(\hat{\alpha}_i)$ given the choice of strategy σ_i. From an outsiders perspective, the expected pay-off for agent i with α_i and $\Theta_i = t$ playing σ_i given Π over all possible realizations of $\hat{\alpha}$ is,

$$V_i(\sigma_i) = \alpha\Phi_{\alpha,t,\sigma_i,\Pi} + \gamma(1 - \Phi_{\alpha,t,\sigma_i,\Pi}) \tag{7}$$

with $\sigma_i \in \{A, B, R\}$. Individual optimality can then be defined as follows.

DEFINITION 1 (Individual Optimality): *The strategy σ_i^* is optimal for the agent from an individual perspective when,*

$$\sigma_i^* = \underset{\sigma_i}{\operatorname{argmax}}\, V_i(\sigma_i)$$

Individual optimality means that an agent uses her social identity cue to maximize her expected pay-off on average over all possible realizations of $\hat{\alpha}_i$. The optimal strategy σ^* is therefore determined by an agent's fitness with respect to a certain task given her type (α, Θ) and the social context Π. I assume agents can compare $V_i(R), V_i(A)$ and $V_i(B)$ and choose their strategy σ_i according to Definition 1. This can be motivated with either a story where agents learn their optimal strategy from their own experience with similar tasks through life, or with a two-selves model like Benabou and Tirole (2006) where a sophisticated self knows the true type, but commits to a strategy σ^* to 'tie the hands' of a less sophisticated 'in-the-moment' self.[8]

[8]See Liqui Lung (2022) for a more elaborate discussion of this assumption.

Population Equilibrium - Let σ be the collection of σ_i. Because N is arbitrarily large, each collection of strategies σ and social context Π generate choices and successes that in turn generate public data $\tilde{\Pi}$ such that,

$$\tilde{\pi}_t(\sigma,\Pi) = \frac{p_t \int \alpha \Phi_{\alpha,t,\sigma,\Pi} f(\alpha) d\alpha}{\sum_{t\in T} p_t \int \alpha \Phi_{\alpha,t,\sigma,\Pi} f(\alpha) d\alpha} \tag{8}$$

where $f(\alpha)$ is the probability density function of α and $\tilde{\pi}_t(\sigma,\Pi)$ is the social identity cue induced by strategies σ and a social context Π. An equilibrium in the model can now be defined as follows.

DEFINITION 2 (Population Equilibrium): *A pair of strategies and a social context* $\{\sigma,\Pi\}$ *constitutes a population equilibrium, when* $\sigma = \sigma^*$ *for all agents given* Π, *and when* Π *is such that,*

$$\Pi = \widetilde{\Pi}(\sigma,\Pi) \tag{9}$$

3 Results

3.1 Individual Choice behaviour

Assuming people choose their optimal strategy according to Definition 1, the model provides insights regarding how and why people focus on certain aspects of their social identity as a function of their type and social context. The following example aims to illustrates this.

Example - Consider a firm in which a priori identical agents choose whether to pursue a career in management (C) or a clerical job (NC). They observe the current pool of successful managers in which women and people from an underrepresented ethnic minority (URM) are underrepresented. Let gender be represented by $\theta^A \in \{1,0\}$, where $\theta^A = 1$ represents being male. Let not belonging to an URM be represented by $\theta^B = 1$, where $\theta^B \in \{1,0\}$. Furthermore, for simplicity, assume $\eta_1^A = \eta_1^B$. Figure 1 shows the thresholds γ_i that follow from the strategies $\sigma_i \in \{A,B,R\}$ for a male URM agent. The arrows show the probabilities $\Phi_{\alpha,t,\sigma_i,\Pi}$ with which the agent chooses to leadership career induced by each strategy σ_i given the social context Π.

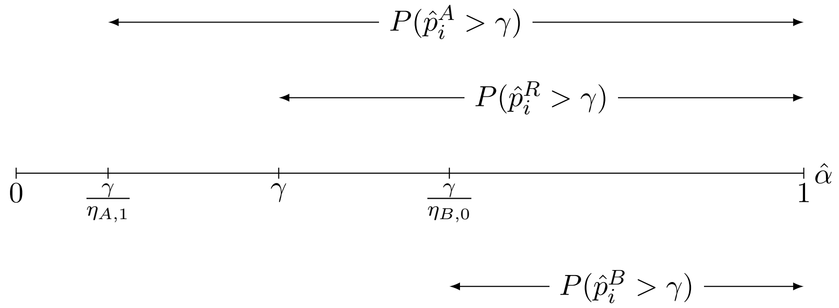

Figure 1: The probabilities to choose $a = C$ induced by $\sigma_i \in \{A, B, R\}$ for an agent with type $t = 10$

Consider first an URM male agent with $\alpha > \gamma$. To maximize expected utility, this agent should choose the leadership career. When he focusses only on the statistic regarding successful URM managers, he will inflate the threshold above which he believes he is good enough to choose the leadership career. This decreases the likelihood he chooses his welfare-maximizing option compared to when he plays $\sigma_i = R$. When the agent focusses on all successful male managers, he deflates the threshold above which he chooses his welfare-maximizing option, which increases the likelihood he chooses the leadership career. To maximize the likelihood he undertakes his welfare-maximzing task over all possible realizations of his noisy belief $\hat{\alpha}_i$, an URM male agent with $\alpha > \gamma$ should therefore believe that the relative overrepresentation of male managers increases his own chances of success in leadership, while he should disregard the fact that URM managers are underrepresented among those currently successful. This is exactly vice versa for an URM male agent with $\alpha < \gamma$. He should believe that belonging to an URM decreases his chances of success in leadership, and disregard the fact that men are overrepresented among the currently successful managers.

Define $\overline{\eta}_t = \max_k \eta_{k,x}$ and $\underline{\eta}_t = \min_k \eta_{k,x}$. Similarly, let $\overline{\kappa}_t = \text{argmax}_k \, \eta_{k,x}$ and $\underline{\kappa}_t = \text{argmin}_k \, \eta_{k,x}$. Then, these results can be generalized as follows.

PROPOSITION 1 (Individually Optimal Belief Formation): *For all agents with $\alpha_i > \gamma$, $\sigma_i^* = \overline{\kappa}_t$ if and only if $\overline{\eta}_t > 1$. For agents with $\alpha_i < \gamma$, $\sigma_i^* = \underline{\kappa}_t$ if and only if $\underline{\eta}_t < 1$. Otherwise, $\sigma_i^* = R$.*

Proposition 1 shows how, in a specific choice setting, agents endogenously determine which dimensions of social identity affect belief formation as a function of their exogenously specified social type, their social context and their underlying ability. This result can give us a motivation for why certain dimensions of social identity become salient to agents in a particular social and choice context.

3.2 Aggregate Choice behaviour

3.2.1 Potential to Improve Decision Making and Inequality

The noise in agents' perceptions can induce two types of errors. A type I error occurs when they choose the *Competence-Driven* task while $\alpha < \gamma$. A type II error occurs when they choose the *Non-Competence-Driven* task while $\alpha > \gamma$. Whether agents are able to reduce the likelihood of committing these error depends on the direction of the bias induced by their possible strategies σ_i given the current social context Π.

Example - Consider again the agents from the previous example in a context where women ($\theta^A = 0$) and URM ($\theta^B = 0$) agents are relatively underrepresented among the successful managers. Table 1 shows for each social type whether they are able to reduce the likelihood of committing a type I respectively type II error.

t	Type I error	Type II error
11	No	Yes
10	Yes	Yes
01	Yes	Yes
00	Yes	No

Table 1: The potential to improve decision making for each social type $\Theta_i = t$

Table 1 shows tnon-URM male agents can only decrease the likelihood of making a type II error, while URM female agents can only decrease the likelihood of making a type I error. Non-URM female agents and URM male agents, on the other hand, can decrease the likelihood of making both types of mistake. These agents are therefore on average more likely to choose their welfare-maximizing task.

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 behaviour, and are defined in Definition 4.

DEFINITION 4: *For a given social context* Π *we define two categories of social types* Θ*:*

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

In other words, an agent with a *mixed* social type belongs to the socially more successful group according to one observable characteristic, but to the socially less successful group according to the other characteristic. An agent with a *one-sided* social type belongs to either the socially more or socially less successful group according to both observable characteristics. Using this definition, we can show one of the main insights of the paper, which is captured in Proposition 2.

PROPOSITION 2 (Potential to Improve Decision Making): *Asymmetry* $\pi_x^k \neq p_x^k$ *along the lines of both observable characteristics* θ^A *and* θ^B *leads to inequalities in the potential to improve decision making across the different social types* $\Theta = t$ *with* $t \in T$*. Specifically, agents with mixed social types can decrease the likelihood of making both types of errors, while agents with one-sided social types can only decrease the likelihood of making one type of error. This induces an on average higher expected utility for agents with mixed social types than one-sided social types.*

Proposition 2 shows how the multidimensionality of social identity only reinforces the potential to improve decision making for agents with *mixed* social types, while it provides no extra benefits for agents with *one-sided* social types. The multi-dimensionality of social types disadvantages agents with *one-sided* social types relative to agents with a *mixed* social type, independent of whether the respective groups are relatively over- or underrepresented among the successful individuals. This result goes therefore against the general intuition that the disadvantages of underrepresentation add up across the different dimensions of social identity. 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 improve decision making on average. Intersectionality in this model can increase the flexibility agents have to use this tool to their advantage and 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 maybe counterintuitive result, we will see that the dynamics this behaviour induces at the aggregate level are in line with what we see in the data.

3.2.2 The Dynamics in Aggregate Choice behaviour

To analyze the dynamics of choice behaviour at the aggregate level, we can aggregate individual choices in two different ways.

DEFINITION 3: *We can analyze the effect of social context on decision making at the aggregate level using the following approaches:*

- **One-dimensional lens**: *aggregating data by creating subgroups defined along the lines of one observable characteristic* θ^k
- **Intersectional lens**: *aggregating data by creating subgroups defined along the lines of social types* t

In the following, I show how using an *Intersectional lens* can shed light on dynamics and inequalities that are not visible with a *One-dimensional lens.*

Example - When we use a *One-Dimensional lens* along the lines of gender, we obtain the potential to improve decision making of male agents by aggregating this potential of URM and non-URM male agents. Similar for female agents. This aggregation shows that, on average, male and female agents have a similar potential to correct for the possible mistakes. There is nevertheless an asymmetry. On average, there are more male agents that are potentially able to correct for a type II error, while there are on average more female agents that are potentially able to correct for a type I error.

An analysis with a *One-Dimensional lens* shows therefore no inequality across subgroups in the potential to improve decision making, only an asymmetry. Only with an *Intersectional lens* we are able to see that this asymmetry is driven by agents with *one-sided* social types, and that not all male and female agents have an equal potential to improve decision making.

Proposition 1 induces probabilities $\Phi_{\alpha,t,\sigma_i^*,\Pi}$ with which agents choose the *Competence-Driven* task. In the following, I illustrate how differences in these probabilities across agents with different types (α, t) drive differences in choice behaviour and average success rates across social groups.

Example - Let $s_i \in \{0, 1\}$, where $s_i = 1$ when $\alpha_i > \gamma$. Let $\hat{p}^{\sigma_i}(s_i, t_i)$ be the belief of an agent with (s_i, t_i) playing σ_i. Figure 2 shows the induced probabilities with which agents choose the leadership career. All male agents and non-URM agents with $\alpha > \gamma$ focus on the social identity cue based on the representation of these respective groups among the successful managers. They therefore choose this career for all realizations of $\hat{\alpha} > \frac{\gamma}{\eta_{k,1}}$ with $k \in \{A, B\}$. Their probability of entering the competition is represented by the top arrows. Agents that are both non-URM and male are indifferent between using π_1^A or π_1^B. URM female agents with $\alpha > \gamma$ and non-URM male agents with $\alpha < \gamma$ choose to *Repress* the use of social identity cues. Their induced probabilities to choose the leadership career are represented by the middle two arrows. Finally, URM agents and female agents with $\alpha < \gamma$ focus on the social identity cue based on the representation of their respective groups among the successful managers. Their probability of entering the competition is represented by the lower two arrows. Agents that are both URM and female are indifferent between using π_0^A and π_0^B.

When we analyze this choice behaviour with a *One-Dimensional lens*, figure 2 shows that male agents have on average a larger probability to choose the leadership career than female agents. This is driven by the fact that all male agents can decrease the likelihood of making a type II error, while all female agents are able to decrease the likelihood of making a type I error. At the same time, Figure 2 shows that female agents choose the leadership career for on average higher values of $\hat{\alpha}$ than male agents. Because $\hat{\alpha}$ is unbiased, this implies that, conditional on choosing this career, female agents have on average a higher success rate than male agents. Because α and the observable characteristics are independently distributed over the population this will not reverse the order on the social identity cues π_1^k and π_0^k. This result is presented in Corollary 1.1 and is similar to the result in Liqui Lung (2022).

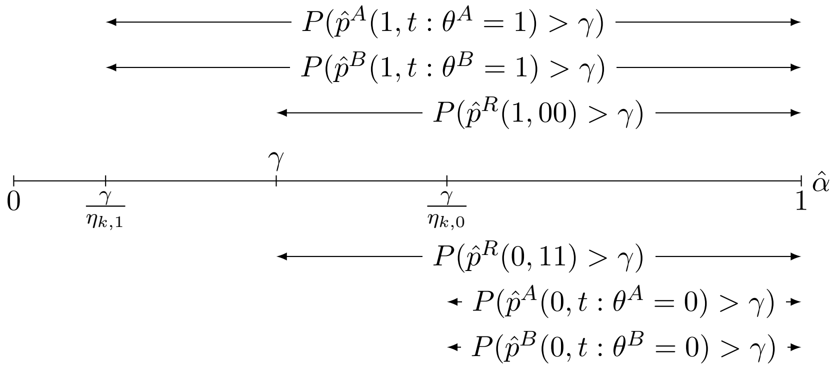

Figure 2: The induced probabilities for agents to choose $a = C$ implied by Proposition 1

COROLLARY 1.1 (One-Dimensional Lens): *For any observable characteristic* θ^k *such that* $\pi_x^k > p_x^k$*, we have 'population effect', such that* $\Phi_{\alpha,t:\theta^k=x,\sigma_i^*,\Pi} > \Phi_{\alpha,t:\theta^k=x',\sigma_i^*,\Pi}$ *and a 'selection effect', such that* $E(\alpha|a = C, t : \theta^k = x) < E(\alpha|a = C, t : \theta^k = x')$*. These effects are such that the order on* π_x^k *and* $\pi_{x'}^k$ *will not be reversed for any* $k \in \{A, B\}$*.*

With two-dimensional social types, an analysis with an *Intersectional lens* sheds nevertheless light on what drives these *one-dimensional* selection and population effects.

Example - Figure 2 shows that non-URM male agents have on average the largest induced probability to choose the leadership career. This is driven by the fact that they are most likely to make a Type II error. URM female agents have on average the smallest probability to choose the leadership career, since they are most likely to make a Type I error. The inequality in the potential of decision making across *mixed* and *one-sided* social types creates *social type-specific* population effects, such that

$$\Phi_{\alpha,11,\sigma_i^*,\Pi} > \Phi_{\alpha,10,\sigma_i^*,\Pi} = \Phi_{\alpha,01,\sigma_i^*,\Pi} > \Phi_{\alpha,00,\sigma_i^*,\Pi} \tag{10}$$

and *social type-specific* selection effects, such that

$$E(\alpha|a = C, t = 11) < E(\alpha|a = C, t \in \{10, 01\}) < E(\alpha|a = C, t = 00) \qquad (11)$$

This shows there are no differences in choice behaviour across agents with a *mixed* social type[9], and one-dimensional selection and population effects are driven by agents with *one-sided* social types.

More generally, let $\tilde{t}_x$ be the *one-sided* social type that has $\theta^k = x$ for all $k \in \{A, B\}$, while $\tilde{t}_{x'}$ is the *one-sided* social type that has $\theta^k = x'$ for all $k \in \{A, B\}$. Let t_{mixed} be any *mixed* social type. Then, the *social type-specific* population and selection effects can be generalized as follows.

COROLLARY 1.2: *Let $\pi_x^k > p_x^k$ for all $k \in \{A, B\}$. We have a social type-specific population effect, such that $\Phi_{\alpha,\tilde{t}_x,\sigma_i^*,\Pi} > \Phi_{\alpha,t_{mixed},\sigma_i^*,\Pi} > \Phi_{\alpha,\tilde{t}_{x'},\sigma_i^*,\Pi}$, and a social type-specific selection effect, such that $E(\alpha|a = C, \tilde{t}_x) < E(\alpha|a = C, t_{mixed}) < E(\alpha|a = C, \tilde{t}_{x'})$. These effects are such that the order on π_x^k and $\pi_{x'}^k$ will not be reversed for any observable characteristic $k \in \{A, B\}$.*

Corollary 1.1 shows how analyzing data with a *One-Dimensional lens* enables us to explain asymmetries in choice behaviour across agents with a different value of each observable characteristic. An *Intersectional lens* shows nevertheless that these *one-dimensional* population and selection effects are predominantly induced by the behaviour of agents with a *one-sided* social type. This is important information for the development of policy that aims to diminish asymmetries in choice behaviour along the lines of a particular dimension of social identity.

Population Equilibrium - How the population and selection effects affect the possible equilibrium outcomes depends on whether they shrink or increase the differences in the induced social context π_x^k and π_t. I characterize the different possible equilibrium outcomes as follows[10].

[9]There can be differences in choice behaviour across agents with different *mixed* social types when $\eta_x^A \neq \eta_x^B$. This does not invalidate the general results and intuition.

[10]The one-dimensional Asymmetric Regime is considered in Liqui Lung (2022)

DEFINITION 5 (Population Equilibrium): *In a* ***Symmetric Regime*** *the allocation of individuals over tasks is symmetric across subgroups, and* $\pi_t = p_t$ *for all* $t \in T$*. In a* ***Asymmetric Regime of Degree 2*** *the allocation of individuals over tasks is asymmetric along the lines of two observable characteristics, and* $\pi_x^k \neq p_x^k$ *for* $k = \{A, B\}$*.*

We can show a *Symmetric Regime* always exists. Take a social context Π such that $\pi_t = p_t$ for all $t \in T$, meaning no social type is relatively over- or underrepresented among the successful individuals. This implies that $\pi_x^k = p_x^k$ for all $k \in \{A, B\}$ and $x \in X$. Therefore, the strategies $\sigma \in \{A, B, R\}$ are equivalent, no matter an agent's type (α, Θ). Therefore, there will be no differences in the induced choice behaviour across social types, and $\tilde{\pi}_{k,x}(\sigma, \Pi) = p_x^k$ for all $k \in \{A, B\}$. In the following example, I show how and why a *Symmetric Regime* can become unstable.

Example - Consider the case in which agents have an extreme version of the response function η, such that,

$$\eta(\pi, p) = \begin{cases} +\infty & \text{if } \pi > p \\ 1 & \text{if } \pi = p \\ -\infty & \text{if } \pi < p \end{cases}$$

Assume a small change in the social context, such that there are slightly more male and non-URM agents among the successful managers, such that $\pi_1^k > p_1^k$ for $k \in \{A, B\}$. Now, with the extreme response function η, all male and non-URM agents with $\alpha > \gamma$ will choose the leadership career, while all female agents and URM agents with $\alpha < \gamma$ will choose the clerical job. Male non-URM agents with $\alpha < \gamma$ and URM female agents with $\alpha > \gamma$ only choose the career when $\hat{\alpha} > \gamma$. Consequently, $\tilde{\pi}_1^k(\sigma, \Pi) > p_1^k$ for both $k \in \{A, B\}$, while $\tilde{\pi}_0^k(\sigma, \Pi) < p_0^k$, and the *Symmetric Regime* becomes unstable.

Using the extreme response function, we can show that any social identity cue $\tilde{\pi}_1^A(\sigma, \Pi)$ is always bounded from above. Let,

$$S_1^A = p_1^A \int_{\alpha > \gamma} \alpha f(\alpha) d\alpha + p_1^A p_1^B \int_{\alpha < \gamma} \int_{\hat{\alpha} > \gamma} \alpha g_\alpha(\hat{\alpha}) f(\alpha) d\alpha d\hat{\alpha} \qquad (12)$$

and,

$$S_0^A = p_0^A p_0^B \int_{\alpha>\gamma}\int_{\hat{\alpha}>\gamma} \alpha g_\alpha(\hat{\alpha}) f(\alpha) d\alpha d\hat{\alpha} + p_0^A p_1^B \int_{\alpha>\gamma} \alpha f(\alpha) d\alpha \tag{13}$$

Then,

$$\frac{\tilde{\pi}_1^A(\sigma,\Pi)}{\tilde{\pi}_0^A(\sigma,\Pi)} \leq \frac{S_1^A}{S_0^A} \tag{14}$$

The analysis for θ^B will be similar. Because $S_1^k > S_0^k$ for both k, the 'extreme' response function gives rise to a stable *Asymmetric Regime of degree 2*. Furthermore, since $S_0^k > 0$ for all k, it follows that $\tilde{\pi}_1^k(\sigma,\Pi) < 1$ for $k \in \{A,B\}$, meaning there will always be a positive number of URM and female agents among the successful individuals. Finally, a larger share of the population with *mixed* social types decreases the difference between S_1^k and S_0^k, which restates how it are especially agents with *one-sided* social types that drive differences in behaviour in this model. Proposition 3 generalizes these insights.

PROPOSITION 3: *An* ***Asymmetric Regime of degree 2*** *in which WLOG* $\pi_x^k > p_x^k$ *for all* $k \in \{A,B\}$ *can co-exist with a* ***Symmetric Regime****. In any* ***Asymmetric Regime****, the order on the social identity cues must be such that,*

$$\pi_{\tilde{t}_{x'}} = \min_{k,x} \pi_x^k \qquad\qquad \pi_{\tilde{t}_x} = \max_{k,x} \pi_x^k$$

where $\tilde{t}_x$ *is such that* $\theta^k = x$*, while* $\tilde{t}_{x'}$ *is such that* $\theta^k = x'$ *for all* $k \in \{A,B\}$*.*

Agents with a *one-sided* social type belonging to the socially more successful groups will be most overrepresented among the successful individuals, because they are most likely to make a type II mistake. Agents with a *one-sided* social type belonging to the socially less successful groups will be most underrepresented among the successful individuals, because they are most likely to make a type I mistake. An *Intersectional lens* sheds light on how asymmetries along the lines of various observable characteristics interact and can induce persistent differences in choice behaviour and representation across social types.

Welfare - As in Liqui Lung (2022), we can define welfare as the aggregate expected utility over all agents in the society. When this is the case, an *Asymmetric Regime* is a Pareto improvement over a *Symmetric Regime*. The intuition behind this result is that in an *Asymmetric Regime*, only those agent that can improve decision making on average with the *social identity cues* change their behaviour. The agents that cannot use social context are not made worse off. Whether an *Asymmetric Regime of Degree 2* leads to an increase in welfare over a *Asymmetric Regime of Degree 1* depends on the following trade-off. On the one hand, when agents have access to multiple social identity cues, agents with a *mixed* social type can decrease the likelihood of making both a Type I and Type II error. This means that, on average, the set of agents that can potentially improve decision making increases. This has a positive effect on welfare. On the other hand, any persistent asymmetry along the lines of an observable characteristic is driven by the asymmetry in the types of error subgroups can potentially correct for. In an *Asymmetric Regime of Degree 1*, this asymmetry is driven by the behaviour of all agents. In an *Asymmetric Regime of Degree 2*, this asymmetry is only driven by the agents with a *one-sided* social type. In equilibrium, we therefore observe smaller deviations of π_x from p_x in the latter than in the former. These smaller deviations lead to smaller effects of social identity cues on decision making, which has a negative effect on welfare.

3.3 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 in belief formation. This changes the strategy set to $\sigma_i \in \{A, B, F, R\}$, where F refers to the strategy in which agents use the *two-dimensional* social identity cue derived from their full social type Θ_i. When $\sigma_i = F$, the corresponding belief $\hat{p}_i^F = \eta_t \hat{\alpha}_i$, where, $\eta_t = \eta(\pi_t, p_t)$. The function $\eta(\pi_t, p_t)$ can be different from $\eta(\pi_x^k, p_x^k)$. People could for example react stronger to *two-dimensional* social identity cues than *one-dimensional* social identity cues. Consequently, let

$$\overline{\overline{\eta}}_t = \max_{t,k \in \{A,B\}} \eta_{k,x}, \eta_t \text{ and } \underline{\underline{\eta}}_t = \min_{t,k \in \{A,B\}} \eta_{k,x}, \eta_t$$

Similarly, define

$$\overline{\kappa}_t = \underset{t,k\in\{A,B\}}{\text{argmax}}\, \eta_{k,x}, \eta_t \text{ and } \underline{\kappa}_t = \underset{t,k\in\{A,B\}}{\text{argmin}}\, \eta_{k,x}, \eta_t$$

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

COROLLARY 2: *For agents with $\alpha > \gamma$, $\sigma_i^* = \overline{\kappa}_t$ if and only if $\overline{\overline{\eta}}_t > 1$. When agents have $\alpha < \gamma$, then $\sigma_i^* = \underline{\kappa}_t$ if and only if $\underline{\underline{\eta}}_t < 1$. Otherwise, $\sigma_i^* = R$.*

Corollary 2 shows that agents will only use the option $\sigma_i = F$, when η_t provides a stronger bias in the direction of their welfare-maximizing task than $\eta_{k,x}$ for any k. The extra strategy affects the outcomes at the aggregate level in the following ways. First, when $\eta_{00} < \eta_{k,0}$ for $k \in \{A, B\}$ and $\eta_{11} > \eta_{k,1}$ for $k \in \{A, B\}$, the introduction of the strategy $\sigma_i = F$ increases both *social type-specific* and *one-dimensional* population and selection effects. Secondly, η_t may induce a stronger bias for a similar pair (π, p) than $\eta_{k,x}$. In this case, it may become optimal for agents with a *mixed* social type to choose $\sigma_i = F$. This will create an asymmetry in the degree to which agents with a mixed social type can correct for a certain type of mistake. Definition 6 helps analyze what happens at the aggregate level.

DEFINITION 6: *An observable characteristic θ^k is dominant when $\pi_t > p_t$ for all t such that $\theta^k = x$, while $\pi_t < p_t$ for all t such that $\theta^k = x'$.*

In other words, we call an observable characteristic *dominant*, when agents with a *mixed* and *one-sided* social type with one realization of this observable characteristic are overrepresented, while agents with a *mixed* and *one-sided* social type with the other realization of this observable characteristic are underrepresented. For an observable characteristic that is not dominant, for one realization agents with a *mixed* social type are overrepresented, while agents with a *one-sided* social type are underrepresented, and for the other realization vice versa. When agents with a *mixed* social type choose $\sigma_i = F$, this will increase the population and selection effects along the lines of the *dominant* observable characteristic, while it will decrease the population and selection effects along the lines of the characteristic that is not dominant. The net effect of the introduction of the strategy F depends on the total effect of this strategy on the behaviour of agents with both *mixed* and *one-sided* social types, and it provides the following Corollary to Proposition 3.

COROLLARY 3: *The introduction of the strategy $\sigma_i = F$ does not invalidate the existence of a Asymmetric Regime of Degree 2.*

3.4 Correlated Observable Characteristics

To simplify the discussion of the multidimensionality of social identity, I assumed the individual observable characteristics are independently distributed over the population. Although this is a reasonable assumption to make in some cases, there are also cases in which observable characteristics are correlated. When this is the case,

$$p_t \neq p_x^A p_x^B$$

In most cases, two social identities that both imply belonging to the socially less or more successful group are positively correlated. For example, belonging to an underrepresented minority group is often positively correlated with belonging to a lower income class. When this type of correlation exists, there will be relatively more agents with a *one-sided* social type than a *mixed* social type. Because agents with *one-sided* social types mainly drive the *one-dimensional* population and selection effects, this type of correlation leads to an increase in the strength of these effects. When one social identity that implies belonging to the socially less successful group and another that implies belonging to the socially more successful group are positively correlated, this increases the fraction of agents with a *mixed* social type relative to the number of agents with a *one-sided* social type. This decreases the strength of the *one-dimensional* population and selection effects.

4 Social Identity and the Strategy Set

In the previous sections, the social type of each agent is fixed and they can freely ignore certain dimensions of this type. It is nevertheless possible to manipulate a 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 that can only be received when belonging 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, economic and social opportunities for whites relative to blacks. Cassan (2015) similarly shows how the Punjab alienation of land act led to a movement of identity-manipulation. This ability to self-determine a social type enlarges their strategy set.

Furthermore, the ability to ignore certain dimensions of the social type may be limited 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 observable characteristic in common. For example, Crenshaw (1991) describes how black women are not able to identify with white women, because their experience in society is so different. The framework developed in this paper allows us to study both the ability to self-identify and the effects of social constraints through simple adjustments to the strategy set. In the following, I discuss what these adjustments look like and what their effects are on aggregate choice behaviour.

4.1 Strategy Restrictions

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. I refer to this type of constraint as *Type-Specific Social Identification*. This type of restriction aims to capture that agents cannot identify with other social types, despite having some characteristics in common.

The main take-aways from the analysis are that, first, *Stigmatization* reinforces the population and selection effects in the dimension of the stigmatized observable characteristic, and decreases these effects in the dimension of the non-stigmatized characteristic. Secondly, *Stigmatization* mainly has a negative effect on the potential to improve decision making of agents with a *mixed* social type. When the stigmatized observable characteristic is not *dominant*, the availability of *two-dimensional* social identity cues can partially mitigate these negative effects. Finally, *Type-Specific Social Identification* reinforces the population and selection effects along the lines of the *dominant* observable characteristic. The population and selection effects for the observable characteristic that is not dominant will be small.

4.1.1 Stigmatization

I introduce two versions of *Stigmatization* in the model. In the first version, an entire observable characteristic is stigmatized. I introduce this version in the model in a step-wise manner. 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 the cues π_x^k and π_t. In the second version of stigmatization, only one value of an observable characteristic is stigmatized, meaning that only those agents with that specific value of the characteristic cannot ignore the characteristic, while agents with another value can. I call this model the *Asymmetric Model*.

Two-Stategy Model - In this model, agents only use *one-dimensional* social identity cues. Assume gender is the stigmatized dimension of social identity. Hence, the strategy set is reduced to $\sigma_i \in \{A, R\}$. In this case, there is no difference in the potential to improve decision making between *mixed* and *one-sided* social types. Only agents with *mixed* types are disadvantaged by stigmatization in this model, while agents with *one-sided* types are not effected at all. Furthermore, agents cannot use the dimension of belonging to an URM in decision making. Because the characteristics θ^k and α are independently distributed over the population, there can be no differences in choice behaviour between URM and non-URM agents. On the other hand, the differences in choice behaviour between male and female agents are now driven by the behaviour of agents with both one-sided social types and *mixed* social types. Therefore, stigmatization induces larger population and selection effects in the dimension of gender than a model without stigmatization[11].

The Three-Strategy Model - Agents can now use both *one-dimensional* and *two-dimensional* social identity cues. When gender is stigmatized, the strategy set becomes $\sigma_i \in \{A, F, R\}$. Like in Section 3.3, there are two different settings. In the first setting, gender is *dominant*, and both URM and non-URM female agents are relatively underrepresented among the successful managers. In this case, non-URM female agents with $\alpha > \gamma$ and URM male agents with $\alpha < \gamma$ cannot use $\sigma_i = F$ 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_t instead of π_x^A, 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 behaviour across URM and non-URM agents. Unlike in the *Two-Strategy Model*, the availability of π_t can therefore induce population and selection effects in dimensions of social identity that are not stigmatized.

In the second setting, URM is *dominant*, and both URM female and male agents are relatively underrepresented among the successful managers. Here, agents with a *mixed* social type maintain their ability to potentially correct for both types of mistakes. In this setting, the availability of the cues π_t 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, and the results are similar to those presented in section 3.3.

[11] Like in Section 3.3, there is an asymmetry in the potential to improve decision making across *mixed* social types. Since agents with *mixed* social types now act exactly as agents with *one-sided* social types, this model presents an extreme case. Consequently, the population and selection effects in the dimension of gender are reinforced, while the population and selection effects in the dimension of belonging to an URM completely disappear.

The Asymmetric Model - Assume WLOG that being female is stigmatized. This means that for female agents the strategy set is restricted to $\sigma_i \in \{A, F, R\}$. Male agents, on the other hand, can use the complete strategy set $\sigma_i \in \{A, B, F, R\}$. First, consider the setting in URM is *dominant*. In this setting, non-URM female agents can potentially escape the negative effects of stigmatization by using the cue π_t instead of π_1^B. If it is optimal to choose $\sigma_i = F$ for all agents with a *mixed* social type, stigmatization does not affect the results at the aggregate level. If it is optimal to choose $\sigma_i = k$ for all agents with a *mixed* social type, then stigmatization only negatively affects non-URM female agents with $\alpha > \gamma$. As a result, the potential of male and URM agents to improve decision making slightly increases, while it slightly decreases for female and non-URM agents. 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 belonging to an URM decreases. These effects are similar, but stronger when gender is *dominant*. In this setting, non-URM female agents are no longer able to potentially correct for type I mistake. This creates a difference in the potential to improve decision making between the different *mixed* social types, while the potential of *one-sided* social types is left unchanged.

4.1.2 Type-Specific Social Identification

The second type of restriction implies that agents cannot ignore any dimension of their social type in belief formation. This type of restriction reduces the strategy set to $\sigma_i \in \{F, R\}$. The first insight this restriction provides is it that, 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 agents can potentially correct for a type II error, while all female agents can potentially correct for a type I error. This induces population and selection effects in the dimension of gender that are driven by the behaviour of agents with both *one-sided* and *mixed* social types. The story is different when we evaluate asymmetry along the dimension of belonging to an URM. Non-URM male agents can potentially correct for a type II error, while non-URM female agents can potentially correct for a type I error. For URM male agents and URM female agents, this is exactly the opposite. Hence, there will only be asymmetry along the lines of belonging to an URM, when $\pi_{11} \neq \pi_{10}$ and $\pi_{00} \neq \pi_{01}$. The exact opposite happens when URM is *dominant*. Therefore, compared to the benchmark case, *Type-Specific Social Identification* increases the population and selection effects along the lines of the *dominant* dimension of social identity, while the population and selection effects along the lines of the not dominant observable characteristic will be small at most.

4.2 Strategy Additions

Agents may be able to manipulate their social type to a certain extent. Self-identification may be costly and for some characteristics, like gender, it may be more flexible than for others, like "being a blondie". To guide intuition, consider an extreme case in which agents have full flexibility in choosing their own realization of an observable characteristic. This means we effectively endogenize this dimension of the social type. When agents can pick and choose their social type as they like, they are able to potentially control for both Type I and Type II errors. Hence, the whole population behaves as an agent with a *mixed* social type and there can be no asymmetry in equilibrium along the lines of this observable characteristic. As we restrict the ability of agents to self-identify, or make it costly, we move towards the benchmark case and increase the size of the population that has a *one-sided* social type. Consequently, the less agents are able to self-identify, the stronger the population and selection effects, and hence the asymmetry in choice behaviour along the lines of this observable characteristic in equilibrium. This example provides intuition for why we particularly observe persistent asymmetry in choice behaviour along the lines of observable characteristics that are costly to manipulate, such as gender, race or social class, and not so much along the lines of observable characteristics that are easy to change and adopt, such as hair color or a particular fashion style. Even if such flexible characteristics may drive behaviour for some time, the example above shows why, eventually, this effect will dissipate, forcing agents to focus on the less flexible characteristics of their social type.

5 Affirmative Action and Decision Making

Evaluations of one-dimensional quota show how its long-term effects are sometimes disappointing, and they can increase the underrepresentation of groups not directly targeted by it.[12] In this section, I use the framework developed in the paper to analyze the effects of affirmative action on individual choice behaviour, and show how the mechanism I consider could help explain the findings in the policy evaluations. I consider as example the application process for a STEM university career, and analyze how a one-time quota ensuring a minimum number of seats for female students affects the decision to apply in the next generation.

Let gender be denoted by $\theta^A \in \{\text{M,F}\}$, representing respectively male and female students. Let $\theta^B \in \{\text{N-URM,URM}\}$ represent whether an individual belongs to an underrepresented minority (URM) or not (N-URM). I simplify notation by writing $\eta_x \equiv \eta_{k,x}$. Assume that, traditionally, female students and students belonging to an URM are underrepresented in the STEM career we consider. To simplify the discussion, assume male and N-URM students are overrepresented to the same degree, such that $\eta_{\text{M}} = \eta_{\text{N-URM}}$. Furthermore, assume male N-URM students are most overrepresented, while female URM students are most underrepresented, and gender is *dominant*. The quota could have affected social context in three different ways.

First, it may have affected the representation of URM and N-URM female students equally. Assume students can only use *one-dimensional* social identity cues. The quota leads to an increase in η_{F} and a decrease in η_{M}. This change will only affect the decision making of students with *mixed* social types. Following the decrease in η_{M}, male URM students are now less likely to apply to the career when $\alpha > \gamma$. Similarly, following the increase in η_{F}, female N-URM students are now more likely to apply when $\alpha < \gamma$. When the quota erases all differences between male and female students, these *mixed* social types lose their ability to correct for the respective mistakes. The quota therefore indeed decreases differences in choice behaviour between female and male students, but through a decrease in the number of high-ability URM male students that apply to the career, and an increase in the number of low-ability N-URM female

[12] 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)

students. The quota also increases differences in choice behaviour between N-URM and URM students, and transfers asymmetries from the dimension of gender to the dimension of belonging to an URM. The availability of *two-dimensional* social identity cues can potentially reduce this spillover effect on the choice behaviour of URM versus N-URM students[13], but can also dampen the effect of the quota all together.[14]

When gender is stigmatized, the quota is equally effective, but the spillover effects of the gender quota on the differences in choice behaviour between N-URM and URM students are smaller.[15] On the other hand, when the URM dimension of identity is stigmatized, a gender quota has very little effect.[16] Finally, when students can only identify with others with the same social type, a quota that affects all female students equally is very effective and decreases the differences in choice behaviour across the gender and URM dimensions simultaneously.

In practice, the quota may nevertheless affect the representation of N-URM female students more than the representation of URM female students. For example, Crenshaw (1991) and Yuval-Davis (2006) discuss how policies targeting women disproportionally benefit white women. Consider the extreme case in which the quota only enhances the representation of N-URM female students. As in the previous case, this leads to an increase in η_{F} and a decrease in η_{M}. Therefore, female N-URM students with $\alpha < \gamma$ are more likely to apply, while male URM students with $\alpha > \gamma$ are less likely to do so. The quota now nevertheless also leads to an increase in $\eta_{\text{F,N-URM}}$,

[13]When students can use *two-dimensional* social identity cues, and $\eta_{\text{M,N-URM}} > \eta_{\text{N-URM}}$, while $\eta_{\text{F,URM}} < \eta_{\text{URM}}$, male N-URM and female URM students will use these cues. Because the quota increases $\eta_{\text{F,URM}}$ and reduces $\eta_{\text{M,N-URM}}$, this reduces the population and selection effects in the dimension of belonging to an URM.

[14]When gender is dominant, students with a *mixed* social type could compensate for the increase in η_F and decrease in η_M with *two-dimensional* cues.

[15]Stigmatization mainly affects behaviour of students with a *mixed* social type. Male URM students with $\alpha < \gamma$ cannot use the cue based on URM students to improve decision making, and are more likely to apply. Similarly, female N-URM students with $\alpha > \gamma$ cannot use the cue based on N-URM students to improve decision making, and are less likely to apply. This dampens the spillover effect of the quota on the differences in choice behaviour between N-URM and URM students.

[16]It can only lead to a small decrease in the differences in choice behaviour between male and female students, when $\eta_{\text{M,N-URM}} > \eta_{\text{N-URM}}$, while $\eta_{\text{F,URM}} < \eta_{\text{URM}}$, and male N-URM and female URM students use their *two-dimensional* cues in decision making.

which induces a decrease in all other *two-dimensional* social identity cues. When $\eta_{\text{F,URM}} < \eta_{\text{URM}}$, the decrease in $\eta_{\text{F,URM}}$ causes female URM students with $\alpha < \gamma$ to be less likely to apply. While the quota enhances the representation of female N-URM students among those that apply, it therefore only further decreases the representation of female URM students. This effect is even stronger, when belonging to an URM is stigmatized, or students can only identify with students with the same social type.

Finally, in the previous sections, we saw that differences in choice behaviour along the lines of a single dimension of social identity are mainly driven by agents with *one-sided* social types. We could therefore consider a quota that specifically targets the representation of female URM students. Such a quota would lead to a simultaneous increase in η_{F}, η_{URM} and $\eta_{\text{F,URM}}$, while it leads to a decrease in η_{M}, $\eta_{\text{N-URM}}$ and $\eta_{\text{M,N-URM}}$. This affects the choice behaviour of students with both *mixed* and *one-sided* social types, and leads to a simultaneous decrease in differences in choice behaviour across both male versus female students, and N-URM versus URM students. Such a quota avoids therefore that differences in representation along the lines of one dimension of social identity are transferred to another dimension. Furthermore, targeting students with *one-sided* social types increases the effectiveness of a quota when one dimension of social identity is stigmatized[17]. When students can only use *two-dimensional* cues, this type of quota is nevertheless not the optimal choice. When $\eta_{\text{F,URM}}$ increases, this automatically leads to a decrease in the relative representation of all other social types. This will therefore decrease the probability of all male students to apply, but it also decreases the probability of N-URM female students to apply. A policy that targets all female students equally will be more effective.

The discussion above shows how a gender quota can increase asymmetries along the lines of another dimension of social identity. Furthermore, the effectiveness of the quota will be hindered when other dimensions of social identity are stigmatized. When zero or one characteristic are stigmatized, a quota targeting underrepresented students with a *one-sided* social type is most effective. It is nevertheless important that students have access to *two-dimensional* social identity cues. When students can

[17]A quota targeting students with *one-sided* social types has a similar effect on the social identity cues related to both dimensions of social identity. This type of quota therefore affects the behaviour of students with *mixed* social types, no matter which dimension of social identity is stigmatized.

only identify with others that have the same social type, a quota will be most effective when it targets the representation of all types of female students equally.

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_x. 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.

Finally, under the assumptions made in this model, the effects of a quota are not achieved in the way policy makers may desire. The increase in representation of female students is obtained through a loss of male students with $\alpha > \gamma$, and a gain in female students with $\alpha < \gamma$ that apply. If we nevertheless introduce a quota that imposes an equal number of seats for female and male students, this makes gender an irrelevant dimension for individuals in the decision making process. The resulting loss in welfare that stems from taking away this instrument from students could be avoided by providing data on other dimensions of social identity, for example regarding previous education or personality traits, that can help students to potentially correct for Type I and Type II errors in decision making, but do not contribute to harmful stereotypes and social norms.

6 Conclusion

This paper shows how intersectionality plays a role in individual choice behaviour. People exploit the different dimensions of their social type to help them cope with noise in decision making that they cannot control for. An intuitive way to distinguish between social types is to separate individuals according to whether they belong to the socially more or less successful type in society. The key insight of this paper is nevertheless that, instead, the relevant distinction to make is between *mixed* social types and *one-sided* social types. Agents with *one-sided* social types are relatively disadvantaged, no matter whether they belong to the socially more or less successful group in society, because they are on average more likely to make mistakes in decision making than agents with *mixed* social types. I therefore show how an *intersectional lens* sheds light on inequalities and insights that are invisible when using a *one-dimensional* lens. These insights can explain why *one-dimensional* policy measures that target underrepresentation do not always have the desired effects, and are useful for the development of potentially more effective *multi-dimensional* policy measures.

They way intersectionality affects choice behaviour at the aggregate level depends on the statistics that are available, the constraints agents face and the flexibility they have to determine their social type. I show how a *multi-dimensional* approach to the design of a quota matters, and how a suboptimal design may be ineffective or simply transfer asymmetries from one dimension of social identity to another. Furthermore, I shed light on the undesirable side-effects of affirmative action policy, and discuss when and how informational policies could effectively accompany a quota to increase the representation of underrepresented groups.

References

D. Atkin, E. Colson-Sihra, and M. Shayo. How do we choose our identity? a revealed preference approach using food consumption. *Journal of Political Economy*, 129(4): 1193–1251, 2021.

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.

T. Breda, J. Grenet, M. Monnet, and C. Van Effenterre. Do female role models reduce the gender gap in science? evidence from french high school. Pse working paper halshs-01713068, Paris School of Economics, 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.

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

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

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.

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.

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.

K. Hoff and J.E. Stiglitz. Striving for balance in economics: Towards a theory of the social determination of behavior. *Journal of Economics Behavior and Organization*, 126:25–57, 2016.

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.

C.W. Liqui Lung. On the origin and persistence of identity-driven choice behavior. Janeway Institute Working Papers 2232, Faculty of Economics, University of Cambridge, 2022.

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

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.

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

P. Rebughini. Agency in intersectionality. towards a method for studying the situatedness of action. *Socio*, 15:189–205, 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.

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

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: Proofs

PROPOSITION 1 (Individually Optimal Belief Formation): *Consider an agent with social type Θ_i. For agents with $\alpha > \gamma$, $\sigma_i^* = \overline{\kappa}_t$ if and only if $\overline{\eta}_t > 1$. Otherwise, $\sigma^* = R$. For agents with $\alpha < \gamma$, $\sigma_i^* = \underline{\kappa}_t$ if and only if $\underline{\eta}_t < 1$. Otherwise, $\sigma^* = R$.*

Proof. Agents choose σ_i to maximize V_i over all possible realizations of $\hat{\alpha}_i$. Consider first agents with $\alpha > \gamma$. The welfare-maximizing choice for these agents is $a = C$. $V_i(\sigma_i) > V_i(R)$ for some $\sigma_i \neq R$ if and only if $\Phi_{\alpha,t,\sigma_i,\Pi} \geq \Phi_{\alpha,t,R,\Pi}$ for some $\sigma_i \in \{A, B\}$. Since $\Phi_{\alpha,t,\sigma,\Pi} = P(\hat{\alpha} > \gamma^{\sigma_i}|\alpha)$, this is the case when $\gamma^{\sigma_i} < \gamma^{R_i}$. This is true if and only if $\pi_x^k \geq p_x^k$ for some $k \in \{A, B\}$. When multiple social identity cues satisfy this condition, agents maximize V_i by choosing σ_i to maximize $\gamma - \gamma^{\sigma_i}$. This is the case when they choose $\sigma_i = \overline{\kappa}_t$. When none of the social identity cues satisfy the condition, they should choose $\sigma_i = R$. Vice versa for agents with $\alpha < \gamma$,$V_i(\sigma_i) > V_i(R)$ if and only if $\Phi_{\alpha,t,\sigma_i,\Pi} \leq \Phi_{\alpha,t,R,\Pi}$ for some $\sigma_i \in \{A, B\}$. This is the case if and only if $\gamma^{\sigma_i} > \gamma^R$, meaning that we need $\pi_x^k \leq p_x^k$ for some $k \in \{A, B\}$. When multiple social identity cues satisfy this condition, agents maximize V_i by choosing σ_i to maximize $\gamma^{\sigma_i} - \gamma$. This is the case when they choose $\sigma_i = \underline{\kappa}_t$. Otherwise, they should *Repress*. □

PROPOSITION 2 (Potential to Improve Decision Making): *Asymmetry $\pi_x^k \neq p_x^k$ along the lines of both observable characteristics θ^A and θ^B leads to inequalities in the potential to improve decision making across the different social types $\Theta = t$ with $t \in T$. Specifically, agents with mixed social types can decrease the likelihood of making both types of error, while agents with one-sided social types can only decrease the likelihood of making one type of error. This induces an on average higher expected utility for agents with mixed social types than one-side social types.*

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 $V(\sigma^*) > V(R)$ when $\sigma^* \neq R$, agents for whom it is optimal to not repress have a higher expected utility. Agents with a *mixed* social type have $V(\sigma^*) > V(R)$ both when $\alpha < \gamma$ and when $\alpha > \gamma$. Agents with a *one-sided* type always have $V(\sigma^*) = V(R)$ either when $\alpha > \gamma$ or when $\alpha < \gamma$. Therefore, when we aggregate all agents with $\alpha > \gamma$ and $\alpha < \gamma$, agents with a *mixed* social type have on average a higher expected utility than agents with a *one-sided* social type. □

PROPOSITION 3: *An* ***Asymmetric Regime of degree 2*** *in which WLOG* $\pi_x^k > p_x^k$ *for all* $k \in \{A, B\}$ *can co-exist with a* ***Symmetric Regime****. In any* ***Asymmetric Regime****, the order on the social identity cues must be such that,*

$$\pi_{\tilde{t}_{\theta'}} = \min_{k,\theta} \pi_x^k \qquad\qquad \pi_{\tilde{t}_\theta} = \max_{k,\theta} \pi_x^k$$

where $\tilde{t}_\theta$ *is such that* $\theta^k = \theta$*, while* $\tilde{t}_{\theta'}$ *is such that* $\theta^k = \theta'$ *for all* $k \in \{A, B\}$*.*

Proof. The social identity cues $\tilde{\pi}_1^k(\Pi, \sigma)$ induced by strategies σ and a social context Π for $k \in \{A, B\}$ are given by,

$$\tilde{\pi}_1^A(\Pi, \sigma) = \frac{S_{11} + S_{10}}{S_{11} + S_{10} + S_{00} + S_{01}} \tag{15}$$

$$\tilde{\pi}_1^B(\Pi, \sigma) = \frac{S_{11} + S_{01}}{S_{11} + S_{10} + S_{00} + S_{01}} \tag{16}$$

with $S_{x^A x^B} = p_\theta^A p_\theta^B \int \alpha \Phi_{\alpha,t,\sigma,\Pi} f(\alpha) d\alpha$ denoting the number of successful agents in the *Competence-Driven* task with social type $t = (x^A, x^B)$. We infer a *Symmetric Regime* always exists. When $\pi_t = p_t$ for all $t \in T$, then, because the observable characteristics θ^A and θ^B are independently distributed, $\pi_1^k = p_1^k$ for all $k \in \{A, B\}$. Then, all strategies $\sigma \in \{A, B, R\}$ are equivalent. Since α and Θ are independent, $\tilde{\pi}_1^k(\Pi, \sigma) = p_1^k$ for all $k \in \{A, B\}$.

Now consider a perturbation of a *Symmetric Regime* such that $\pi_{11}^\delta = \pi_{11} + \delta$, while $\pi_{00}^\delta = \pi_{00} - \delta$. Consequently, $\pi_{11}^\delta > \pi_{00}^\delta$, while $\pi_{10}^\delta = \pi_{01}^\delta$. From Equations (15) and (16), we can infer that such a perturbation has a symmetric effect on the induced social identity cues $\tilde{\pi}_1^k(\Pi, \sigma)$ for $k \in \{A, B\}$. Let $\mathcal{S}^{sym} = \{\Pi : \pi_{11} = 1 - \pi_{00} \text{ and } \pi_{10} = \pi_{01}\}$ be the set of all social contexts Π in which the induced social identity cues $\tilde{\pi}_1^k(\Pi, \sigma)$ are symmetric for $k \in \{A, B\}$. Similarly, we can now define the set $\mathcal{S}_\delta^{sym} = \{\Pi : \pi_{11} = 1 - \pi_{00}, \pi_{10} = \pi_{01} \text{ and } \pi_{11} + \pi_{10} > \frac{1}{2}\}$, that contains all Π that are induced by a perturbation δ such that $\pi_{11}^\delta = \pi_{11} + \delta$, while $\pi_{00}^\delta = \pi_{00} - \delta$. It follows that $\mathcal{S}_\delta^{sym} \subset \mathcal{S}^{sym}$. Finally, let $\tilde{\Pi}(\sigma, \Pi_\theta)$ be a continuous function.

We now have a non-empty, compact and convex set $\mathcal{S}_{\delta}^{sym}$, and a continuous function $\tilde{\Pi}(\sigma, \cdot) : \mathcal{S}_{\delta}^{sym} \rightarrow \mathcal{S}_{\delta}^{sym}$.Therefore, following Brouwer's fixed point theorem, there exists a fixed point $\Pi^* \in \mathcal{S}_{\delta}^{sym}$ such that,

$$\Pi^* = \tilde{\Pi}(\mathcal{S}_{\delta}^{sym})$$

Finally, because $\Pi^* \in \mathcal{S}_{\delta}^{sym}$ we will have $\pi_{11}^* > \pi_{10}^* = \pi_{01}^* > \pi_{00}^*$. Therefore, $\pi_1^k > \pi_0^k$ for all $k \in \{A, B\}$ and an *Asymmetric Regime of Degree 2* exists. That $\pi_{11}^* > \pi_{10}^* = \pi_{01}^* > \pi_{00}^*$ can be the only order that can exist in such a regime follows from Corollary 1.2. □

COROLLARY 1.1 (One-Dimensional Lens): *For any observable characteristic θ^k such that $\pi_x^k > p_x^k$, we have population effect $\Phi_{\alpha,t:\theta^k=x,\sigma_i^*,\Pi} > \Phi_{\alpha,t:\theta^k=x',\sigma_i^*,\Pi}$ and a selection effect $E(\alpha|a = C, t : \theta^k = x) < E(\alpha|a = C, t : \theta^k = x')$. These effects are such that the order on π_x^k and $\pi_{x'}^k$ will not be reversed for any $k \in \{A, B\}$.*

Proof. Assume WLOG that $\pi_1^k > p_0^k$. Then, all agents with $\alpha > \gamma$, $\theta^k = 1$ and $k = \overline{\kappa}_t$, and all agents with $\alpha < \gamma$, $\theta^k = 0$ and $k = \underline{\kappa}_t$ will choose $\sigma = k$. Because the observable characteristics are independently distributed over the population,

$$\Phi_{\alpha>\gamma,t:\theta^k=1,\sigma,\Pi} > \Phi_{\alpha>\gamma,t:\theta^k=0,\sigma,\Pi} \qquad \text{and} \qquad \Phi_{\alpha<\gamma,t:\theta^k=1,\sigma,\Pi} > \Phi_{\alpha<\gamma,t:\theta^k=0,\sigma,\Pi} \tag{17}$$

Therefore,

$$\Phi_{\alpha,t:\theta^k=1,\sigma,\Pi} > \Phi_{\alpha,t:\theta^k=0,\sigma,\Pi}$$

Because N is arbitrarily large, this translates into population fractions.

Let $\gamma_\theta = \frac{\gamma}{\eta(\pi_x^k, p_x^k)}$. Because $\gamma_1 < \gamma$, while $\gamma_0 > \gamma$, agents with $\theta^k = 1$ choose the *Competence-Driven* task on average for lower values of $\hat{\alpha}$ than agents with $\theta^k = 0$. Because $\hat{\alpha}$ is unbiased, this implies,

$$E(\alpha|a = C, t : \theta^k = 1) < E(\alpha|a = C, t : \theta^k = 0)$$

The number of expected successful agents is of type (α, t) is given by $P(a = C|\alpha, t)E(\alpha|a = C, t)$. Because the observable characteristics are independently distributed over the population from α, the behaviour of agents in absence of asymmetry along the lines of θ^k is symmetric across the group of agents with $\theta^k = 1$ and the group of agents with $\theta^k = 0$. Let $S_{\theta^A\theta^B} = p_x^A p_x^B \int \alpha \Phi_{\alpha,t,\sigma,\Pi} f(\alpha) d\alpha$. To prove that the population and selection effects do not reverse the order on $(\pi_x^k - p_x^k)$ we need to show that, $S_1 > S_0$, where $S_1 = S_{11} + S_{10}$ and $S_0 = S_{00} + S_{01}$. This can be demonstrated by writing,

$$S_1 = p_1^A \int_{\alpha>\gamma} \int_{\hat{\alpha}>\gamma_1} \alpha g_\alpha(\hat{\alpha}) f(\alpha) d\alpha d\hat{\alpha} + \int_{\alpha<\gamma} \int_{\gamma<\hat{\alpha}<\gamma_0} \alpha g_\alpha(\hat{\alpha}) f(\alpha) d\alpha d\hat{\alpha} + p_1^A \int_{\alpha<\gamma} \int_{\hat{\alpha}>\gamma_0} \alpha g_\alpha(\hat{\alpha}) f(\alpha) d\alpha d\hat{\alpha}$$

Since,

$$S_0 = p_0^A \int_{\alpha>\gamma} \int_{\hat{\alpha}>\gamma} \alpha g_\alpha(\hat{\alpha}) f(\alpha) d\alpha + p_0^A \int_{\alpha<\gamma} \int_{\hat{\alpha}>\gamma_0} \alpha g_\alpha(\hat{\alpha}) f(\alpha) d\alpha d\hat{\alpha} + d\hat{\alpha}$$

It follows that when $p_0^A = p_1^A$ and $\gamma_1 < \gamma < \gamma_0$, then $S_1 > S_0$. □

COROLLARY 1.2: *Let $\pi_x^k > p_x^k$ for all $k \in \{A, B\}$. We have a social type-specific population effect, such that $\Phi_{\alpha,\tilde{t}_x,\sigma_i^*,\Pi} > \Phi_{\alpha,t_{mixed},\sigma_i^*,\Pi} > \Phi_{\alpha,\tilde{t}_{x'},\sigma_i^*,\Pi}$, and a social type-specific selection effect, such that $E(\alpha|a = C, \tilde{t}_x) < E(\alpha|a = C, t_{mixed}) < E(\alpha|a = C, \tilde{t}_{x'})$. These effects are such that the order on π_x^k and $\pi_{x'}^k$ will not be reversed for any observable characteristic $k \in \{A, B\}$.*

Proof. Consider first the agents with social type $\tilde{t}_x$. When $\alpha > \gamma$, $\gamma_{\tilde{t}_x} = \max_{t\in T} \gamma_t$. Therefore, $\Phi_{\alpha>\gamma,\tilde{t}_x,\sigma_i^*,\Pi} = \max_{t\in T} \Phi_{\alpha>\gamma,t,\sigma_i^*,\Pi}$. When having $\alpha < \gamma$, they will play $\sigma_i^* = R$, and therefore $\Phi_{\alpha<\gamma,\tilde{t}_x,\sigma_i^*,\Pi} = \max_{t\in T} \Phi_{\alpha<\gamma,t,\sigma_i^*,\Pi}$. Consequently, $\Phi_{\alpha,\tilde{t}_x,\sigma_i^*,\Pi} = \max_{t\in T} \Phi_{\alpha,t,\sigma_i^*,\Pi}$. For agents with social type $\tilde{t}_{x'}$ and $\alpha > \gamma$, $\Phi_{\alpha>\gamma,\tilde{t}_{x'},\sigma_i^*,\Pi} = \min_{t\in T} \Phi_{\alpha>\gamma,t,\sigma_i^*,\Pi}$. When having $\alpha < \gamma$, $\gamma_{\tilde{t}'_x} = \min_{t\in T} \gamma_t$ Therefore, $\Phi_{\alpha<\gamma,\tilde{t}_{x'},\sigma_i^*,\Pi} = \min_{t\in T} \Phi_{\alpha<\gamma,t,\sigma_i^*,\Pi}$. Consequently, $\Phi_{\alpha,\tilde{t}_{x'},\sigma_i^*,\Pi} = \min_{t\in T} \Phi_{\alpha,t,\sigma_i^*,\Pi}$. Agent's with social types t_{mixed} can use their social identity cues to bias their own noisy perception both towards undertaking the task and the outside option. Following Proposition 1, $\Phi_{\alpha>\gamma,t_{mixed},\sigma_i^*,\Pi} \leq \Phi_{\alpha>\gamma,\tilde{t}_x,\sigma_i^*,\Pi}$, and $\Phi_{\alpha<\gamma,t_{mixed},\sigma_i^*,\Pi} \geq \Phi_{\alpha<\gamma,\tilde{t}_{x'},\sigma_i^*,\Pi}$. It follows therefore that $\Phi_{\alpha,\tilde{t}_x,\sigma_i^*,\Pi} > \Phi_{\alpha,t_{mixed},\sigma_i^*,\Pi} > \Phi_{\alpha,\tilde{t}_{x'},\sigma_i^*,\Pi}$. The selection effect follows from these population effects as described in the proof of Corollary 1.1, as well as the proof of the fact that the population and selection effects never reverse the order of representation in the social context. □

COROLLARY 2: *For agents with $\alpha > \gamma$, $\sigma_i^* = \overline{\kappa}_t$ if and only if $\overline{\overline{\eta}}_t > 1$. When agents have $\alpha < \gamma$, then $\sigma_i^* = \underline{\kappa}_t$ if and only if $\underline{\underline{\eta}}_t < 1$. Otherwise, $\sigma_i^* = R$.*

Proof. When agents choose σ_i to maximize $V(\sigma_i)$, it follows that they will choose $\sigma_i = F$ if and only if $F = \text{argmax}_{\sigma_i \in \{A,B,F,R\}} V(\sigma_i)$. When $\alpha > \gamma$, maximizing $V(\sigma_i)$ is equivalent to maximizing $\Phi_{\alpha,t,\sigma_i,\Pi}$. We only have $\Phi_{\alpha,t,F,\Pi} = max_{\sigma_i \in \{A,B,F,R\}} \Phi_{\alpha,t,\sigma_i,\Pi}$ when $\pi_t > p_t$ and $F = \overline{\kappa}_t$. Vice versa, when $\alpha < \gamma$. □

COROLLARY 3: *The introduction of the strategy $\sigma_i = F$ does not invalidate the existence of a Asymmetric Regime of Degree 2.*

Proof. Consider a social context $\Pi \in \mathcal{S}^{sym}$ and perturb this social context, such that $\Pi^\delta \in \mathcal{S}_\delta^{sym}$. If $F = \text{argmax}_{t,k \in \{A,B\}} \eta(\pi_t, p_t), \eta(\pi_x^k, p_x^k)$ for some $t \in \{10, 01\}$, then $\tilde{\pi}_{01} \neq \tilde{\pi}_{10}$ and $\tilde{\Pi}(\sigma, \Pi^\delta) \notin \mathcal{S}_\delta^{sym}$. Therefore, $\tilde{\Pi}(\sigma, \Pi^\delta) \in \mathcal{S}_\delta^{sym}$ if and only if $k = \text{argmax}_{t,k \in \{A,B\}} \eta(\pi_t, p_t), \eta(\pi_x^k, p_x^k)$ for all $t \in \{10, 01\}$. When this condition is met, we can again use Brouwer's fixed point theorem to show that there exists a fixed point Π^* such that $\Pi^* = \tilde{\Pi}(\sigma, \Pi^*)$. □