On the Origin and Persistence of Identity-Driven Choice Behavior

Caroline W. Liqui Lung *

Version: June 1, 2021

This paper provides a novel explanation for identity-driven choice behavior, that does not rely on built-in differences in preferences, discrimination or social pressures, but works through the optimal management of confidence. Agents observe social identity cues that stem from the prevalence of their subgroup among the successful individuals in the social context, and find it rational to let their choices be driven by these social identity cues, even when these cues do not directly affect utility. The influence on choice differs across types, but is not driven by ability differences between subgroups. I show the existence of a stable population equilibrium in which the task allocation and the use of social identity cues differ between a priori identical subgroups.
(JEL: D-81, D-91, I-240, Z-13)

*Paris School of Economics (PSE, ENPC) (Email: caroline.liqui-lung@psemail.eu). I would like to thank Olivier Compte, Philippe Jehiel, Abhijit Banerjee, Muhamet Yildiz, Daron Acemoglu, Glenn Ellison, Sara Ellison, Dilip Mookherjee, Gabrielle Fack, Suanna Oh, Louis-Gaetan Giraudet, Jean-Marc Tallon, Laurent Lamy and the participants of the TOM Seminar (PSE), the MIT Economic Theory Lunch Seminar, the Young Applied Economists Webinar (Warwick/Monash) and the Economic Theory Reading Group (PSE/Sciences Po) for their useful comments and suggestions.

1 Introduction

Examples of identity-driven choice behavior are widely available, and show that this behavior plays an important role in determining people's opportunities in both education and the labor market. Akerlof and Kranton (2000) show that social identity influences people's occupational choice, Nosek et al. (2009) show it can be a better predictor for choice of major than SAT scores, and social identity drives decisions regarding educational investment (Bursztyn and Jensen (2015), Austen-Smith and Fryer (2005)). The economic literature, in the trend of Akerlof and Kranton (2000), models the role of social identity in this type of choice behavior through build-in arguments in the utility function that capture matters as self-image or social pressure, and specifies identity-contingent behavior exogenously by imposing social norms or stereotypes. Although this literature leads to valuable insights, it still leaves some questions unanswered. Because where does this identity-contingent behavior come from and what makes it so persistent? And is this behavior a reflection of the internal manifestations of a certain identity or rather a reflection of the social context? And why do some social identities become salient in a choice context, while others do not? This paper builds on the current literature by addressing these questions in a setting of decision making under uncertainty.

The main novelty of the paper is that it endogenizes the use of social identity in decision making under uncertainty, without assuming social identity directly affects utility. The uncertainty stems from the main assumption I make in the paper, namely that the inference process through which agents learn about their ability is always incomplete. Agents have to create an estimate of the probability of success of a task in which their ability matters. In this process, I give agents the option to use information related to their social identity that stems from what agents observe in the social context. I show that some agents, depending on their type, can use this information as an instrument to bias their beliefs in the correct direction to improve decision making, even when it is informationally irrelevant for decision making. I derive the optimal use of social identity for different types of agents conditional on the social context and show the effects of this behavior at the aggregate level by analyzing the equilibrium in the social context, which consists of mutually stable choices of the use of social identity and tasks.

Agents are described by an ability type and an observable characteristic. The observable characteristic can represent matters such as gender, race, social class or caste and is independently distributed from the ability type. Agents choose between a task with a competitive character, of which the probability of success depends on their ability type, and a task with a non-competitive character, of which the probability of success is known and the same for everybody. They derive utility from being successful. The uncertainty in decision making stems from the fact that agents can only imperfectly learn about their ability type and get a noisy estimate that makes them either overly optimistic or pessimistic about their probability of success of a competitive task. Agents also observe the degree to which their own social group, based on their observable characteristic, prevails among the successful people in the social context. To form an estimate of the probability of success of a competitive task, I let agents choose between two possible models of the world. Agents can either choose a correctly specified model of the world, in which they believe their observable characteristic is irrelevant for their probability of success. In this case, they base their estimate only on the noisy estimate of their ability. Or, agents can choose to hold an incorrectly specified model of the world in which they believe their observable characteristic matters for their probability of success. In this case, they include the information they can retrieve from their social context, which I will call the '*social identity cue*', in their decision-making process.

I show that, although these social identity cues are irrelevant in a Bayesian sense, they become valuable when their use generates a bias towards the welfare-maximizing choice of task. For relatively high ability type agents that are under-confident, biasing the estimate of their chances of success upward is welfare improving, while for relatively low ability type agents that are overconfident, this is the case when biasing the estimate of their chances of success downward. As a result, relatively high ability type agents will wish to use their social identity cue if they belong to a socially more successful group, while they will wish to avoid it when they belong to a socially less successful group. The opposite holds for relatively low ability type agents. This implies that, when the inference process of agents is incomplete, it becomes useful to incorporate signals in the decision function that would otherwise be ignored. Social identity can therefore drive choice behavior, even when it does not directly affect utility. This shows that agents can use their social identity to improve decision making on average[1].

The current economic literature models the effect of social identity on decision making predominantly through the utility derived from self-image (Akerlof and Kranton (2000), Benabou and Tirole (2011)) or through the psychological costs of interaction with others. These psychological costs can stem from the fear of being punished by peers for not complying with social norms (Akerlof and Kranton, 2000), the fear of being rejected

[1]The idea that agents tailor the use of social identity to the relevant social context is consistent with the results of LeBoeuf et al. (2010), who show that agents adjust their social identity to the decision making context, but, at the same time show that not all choices of the participants assimilated to the salient identity, indicating that people sometimes use their social identity in decision making, but sometimes not.

by peers (Austen-Smith and Fryer, 2005), costly interaction with people different from yourself (Battu et al., 2007) or other forms of social pressure[2]. This model provides a different, but possibly complementary view, and shows that the use of social identity can be a mechanism that gives agents the option to manage confidence. Specifically, the use of social identity allows agents to optimally manage the degree of over- or under-confidence regarding their ability through a distinct processing of their noisy estimates. This causes agents to be sometimes too pessimistic, sometimes to be too optimistic about their chances of success[3]. The optimal use of social identity results in the optimal management of confidence to improve decision making. This approach could be extended to situations in which there is a real value to biased confidence, as in Compte and Postlewaite (2004). Furthermore, contrary to Compte and Postlewaite (2004), Hoff and Stiglitz (2010) and Benabou and Tirole (2002), I show that, not only optimism and overconfidence can have instrumental value, but also pessimism and under-confidence.

The options to manage confidence available to agents depend on their observable characteristic, and are therefore asymmetric across different types of agents and will depend on the social context[4]. This asymmetry can potentially lead to different choices of tasks across a priori identical subgroups. This feeds differences in the prevalence of a particular subgroup among the successful people, which in turn fuels the differential use of social identity cues in decision making. I show the existence of a stable population equilibrium in which the task allocation and the use of social identity cues differ between subgroups. I show therefore that, even when agents manage to use their social identity in an optimal way from an individual perspective, asymmetries across a priori identical subgroups can persist. This type of population equilibrium can especially arise when we consider *Competitive* tasks at which only a small fraction of the population succeeds, and always co-exists with a population equilibrium in which the task allocation and the use of social identity cues are the same across subgroups. Furthermore, the degree of asymmetry in the social context across subgroups observed in equilibrium increases when agents react stronger to the social context. Finally, an equilibrium in which a certain subgroup is relatively overrepresented among the successful individuals can exists for a larger set of parameters when this subgroup represents a majority of the population.

[2]See Busztyn and Jensen (2017) for a review.

[3]Evidence for this distinct processing of information can be found in the results of Exley and Kessler (2019), who show that, when women and men receive the same signal about performance related to a test including analytical questions related to math and science, men rate themselves an average of 61 out of 100 while women only rated themselves a 46 out of 100.

[4]This is line with results of experiments that show that, on average, men are more overconfident than women in fields that have a strong male connotation, while the opposite is true in fields that have a strong female connotation (e.g. Coffman (2014) and Flory et al. (2010)).

Interestingly, the use of social identity cues induces both a difference in the propensity to choose the competitive task across subgroups, and a difference in mean competence: for agents belonging to the socially less successful subgroup, choosing the competitive task requires a higher value of the noisy estimate, because these agents cannot use the social identity cue to boost up beliefs. As a result, agents belonging to this subgroup have a lower propensity to choose the competitive task, but, conditional on choosing this task, they tend to be more competent on average than agents belonging to the socially more successful group in the social context[5]. This is in line with the results of Niederle and Vesterlund (2007), who show that too few high skill women and too many low skill men enter competitive math-related tasks. Finally, I show that the use of social identity especially drives the behavior of agents with average ability levels, because agents with extreme ability levels are always more likely to make the welfare-maximizing choice, independent of the social context. This could explain why Buser et al. (2014) find that the gender gap in curriculum choice shows up precisely at the mean: while average men choose highly mathematical curricula, average women choose very humanities-intensive curricula, which causes women to be over-represented in the latter, while men are overrepresented in the former.

The paper shows that social identity cues can have instrumental value in decision making under uncertainty through the optimal management of confidence, and it therefore provides a novel micro-foundation for identity-driven choice behavior, that is embedded in the social context. The paper is different from the current literature that endogenously determines such behavior, such as Benabou and Tirole (2011), Shayo (2009) and Shayo (2020), in that social identity does not affect utility directly, and because I consider decision making under uncertainty, where the uncertainty stems from the incomplete inference process of agents. The equilibrium results provide an explanation for the persistence of this behavior and show how it interacts with the social context. The paper therefore contributes to the literature on stereotypes (e.g. Bordalo et al. (2016)) and the literature on the social determination of behavior (Hoff and Stiglitz, 2016). Finally, the literature on discrimination and affirmative action also shows the existence of a population equilibrium with an asymmetric allocation over tasks of a priori identical subgroups[6], but the novelty of this model is that the results are obtained without assuming any strategical interaction between agents.

[5]A study by S&P market intelligence shows that men outnumber women in the CFO job by about 6.5 to 1. Companies appointing female CFO's saw nevertheless a 6% increase in profits and an 8% better stock return compared to companies appointing male CFO's. Moreover, female CFO's brought in $1.8 trillion of additional cumulative profit and therefore significantly outperformed their male peers.

[6]See Benhabib et al. (2010) for an review.

The paper is organized as follows. Section 2 introduces the model. Section 3 presents the results. Section 4 discusses the assumptions of the model. Section 5 presents applications related to role models, individual feedback, stereotypes, affirmative action and oppositional identities. Section 6 concludes. All proofs can be found in Appendix 2.

2 The Model

I consider a society with $i = 1, ..., N$ agents, with N arbitrarily large. Each agent has an *ability type* that represents for example the agent's leadership qualities, mathematical ability or intelligence. This ability type is unobservable to both the agent herself and all other agents in the society and it is denoted by the continuous variable $\alpha \in [0, 1]$. The ability type is distributed over the population following a distribution function $f(\alpha)$ and fixed for each individual agent, such that each agent has an ability type α_i. Secondly, agents are described by an *observable characteristic* that can represent a wide array of items, such as the agent's gender, ethnicity or social class. The observable characteristic of each agent is public information, meaning that both the agent herself and all other agents in the society can observe it. I let the binary variable $\theta \in \{0, 1\}$ denote this characteristic[7] and p_θ the fraction of the population with observable characteristic θ. Consequently, each agent i has an observable characteristic θ_i and is fully described by her type $\{\alpha_i, \theta_i\}$. To isolate the mechanism through which social identity affects choice in this model, I assume that the ability type and the observable characteristic are independently distributed over the population.[8]

Each agent has to make a choice between two types of tasks. The first type of task has probability of success that is a function of the individual-specific ability type, and has therefore a *Competitive* character. The second type of task can be considered a *Non-Competitive* outside option, and has a probability of success that is fixed, known and the same for all agents. This setting can represent choices such as the choice between a management position and a clerical position and is modelled as follows. Each agent has to take an action $a \in \{C, NC\}$, where $\{C, NC\}$ represent classes of tasks of respectively the *Competitive* and *Non-Competitive* type. The outcome of each task can be either *'success'* or *'failure'* and is represented by the outcome variable $Y_i \in \{1, 0\}$, that is public information.

[7]To simplify the exposition of the model, I use a simple binary variable, but the model can be easily extended to include observable characteristics represented by non-binary variables.

[8]In the Discussion section I show that the results derived from this model are generally robust in a setting where α is correlated with θ.

Let $p_{i,a} = P(Y_i = 1|a)$ be the probability of success for agent i given she has chosen action a. To simplify the exposition of the model, I let the probability of success for *Competitive* tasks be equal to the individual-specific ability type[9]. Therefore,

$$p_{i,C} = \alpha_i \tag{1}$$

The probability of success of *Non-Competitive* tasks is fixed and the same for all agents, and

$$p_{i,NC} = \gamma \tag{2}$$

with $\gamma \in [0, 1]$. More generally, this value γ can be interpreted as the attractiveness of the *Non-Competitive* task relative to the *Competitive* task. Agents derive utility from being successful and the utility function can therefore be represented by $u_i = Y_i$. Each agent chooses her action a to maximize $E(u_i)$. This means that a fully rational agent would choose the *Competitive* task if and only if $p_{i,C} > \gamma$.

Noisy Estimates - Agents are nevertheless only able to imperfectly learn about their ability type α_i. They have access to a noisy estimate $\hat{\alpha}_i$ that is correlated with their ability level α_i and that makes them either overconfident or under-confident. I make the following assumption about this noisy estimate and the inference process it stems from.

ASSUMPTION 1: *The noisy estimate $\hat{\alpha}_i$ stems from an unmodelled inference process that is always incomplete. Specifically, $\hat{\alpha}_i \sim g_{\alpha_i}$, with $E(\hat{\alpha}_i) = \alpha_i$, while $V(\hat{\alpha}_i)$ is independent of α_i.*

The noisy estimate $\hat{\alpha}$ can be considered as the best estimate the agent can obtain given her information set and given the assumption that she is not able to draw all possible inferences given the model as a Bayesian would. The assumption that the inference process is incomplete is supported by a large body of empirical evidence[10]. Another

[9] In general ability may contribute differently to success across different *Competitive* tasks, hence one may expect the relationship between ability and success to be more complex.

[10] Kahneman et al. (1997) show that the incompleteness can be caused by the fact that remembered utility is different from experienced utility and that people make choices based on remembered utility that violate dominance. Similarly, Kahneman and Lovallo (1993) show that people systematically neglect the statistics of the past in evaluating current plans. Attribution theory shows that agents sometimes erroneously ascribe outcomes to external factors and not to their own ability or vice versa (Kelley, 1971).

interpretation that is consistent with this modelling is that what is noisy is the estimation of the probability of success $p_{i,C}$ rather than α, and agents are unable to derive the exact correlation between the task and the degree to which ability matters for that task. Finally, the assumption that the noisy estimate is unbiased can be challenged (Mobius et al., 2014). The main point is nevertheless that a bias in the noisy estimate is not the mechanism that drives the results in this model[11].

Public Data - Besides the noisy estimate $\hat{\alpha}$, agents also have access to public data. This public data consists of the earlier defined variables that are observable to all agents in the society, namely the outcome variables and the observable characteristics of other agents that have made a similar choice at some earlier point in time. Society typically structures public information. An example of this structure is that agents perceive the fraction of successful individuals at task C conditioned on their observable characteristic. For the exposition of the model, I propose the following structure on information, but this specific structure will not drive the results.

Let $\theta_i \in \Theta \equiv \{0,1\}$ and let $\theta \in \Theta$. Let $\mathcal{N}_{C,\theta} = \{i \in N, \theta_i = \theta, a_i = C\}$ be the set of all individuals of type $\theta_i = \theta$ that have chosen the *Competitive* task. Let $\mathcal{N}_C = \{i \in N, a_i = C\}$ be the set of all individuals that have chosen the *Competitive* task, which implies $\mathcal{N}_{C,\theta} \subset \mathcal{N}_C$. Society then provides the statistic,

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

which is the fraction of successful individuals with characteristic θ among all successful individuals that have chosen the *Competitive* task[12]. I call this fraction π_θ the '*social identity cue*' for an agent with observable characteristic θ. The social context of the population is defined as the vector $\Pi = (\pi_\theta)_{\theta \in \{0,1\}}$. Because α and θ are independently distributed over the population, the social context contains no information about the individual-specific probability of success when undertaking a *Competitive* task. Instead, I will introduce the option to agents to bias their noisy estimates $\hat{\alpha}$ with the use of this public data.

Also, agents could have imperfect memory (Piccione and Rubinstein, 1997)

[11]See Discussion section for a discussion on what happens when $V(\hat{\alpha}_i)$ is correlated with α_i.

[12]The fact that I focus only on the successful individuals that have chosen the task and not on all agents that have chosen this task captures the survivors bias (Denrell, 2003). This bias implies that, although agents can observe the outcome variable of other agents, it is more likely that agents are also able to observe the choice of task of other agents when they have successfully completed this task. We usually do not observe those who tried but failed.

From Noisy Estimate to Beliefs/Choice - To choose an action $a \in \{C, NC\}$, agents need an estimate of $p_{i,C}$ to evaluate whether they are 'good enough' to undertake the *Competitive* task. I introduce the following family of belief formation processes with which agents process their noisy estimate $\hat{\alpha}$ into an estimate of the probability of success of a *Competitive* task $\hat{p}_{i,C}$ and I will assume agents have some discretion in finding out which belief formation process suits them best. Specifically, I assume agents have a natural 'urge' to look at others like them when they are not sure what to do or believe and people have to option to either *Repress* or *Not Repress* this urge. Agents will therefore choose a strategy $\sigma_i \in \{R, NR\}$ and,

$$\hat{p}_{i,C} = \begin{cases} \hat{\alpha}_i & \text{if } \sigma_i = R \\ \eta(\pi_{\theta_i} - p_{\theta_i})\hat{\alpha}_i & \text{if } \sigma_i = NR \end{cases} \tag{3}$$

where η is a 'reaction function' that is non-decreasing in $(\pi_\theta - p_\theta)$ such that,

$$\eta(\pi_\theta - p_\theta) = \begin{cases} > 1 & \text{if } \pi_\theta > p_\theta \\ 1 & \text{if } \pi_\theta = p_\theta \\ < 1 & \text{if } \pi_\theta < p_\theta \end{cases} \tag{4}$$

I motivate η and the family of belief formation processes using a subjective prior model in Appendix 1. The choice of strategy σ_i can be interpreted as a choice between two different models of the world. When the agent chooses $\sigma_i = R$, she chooses a correctly specified model of the world in which ability and observable characteristics are uncorrelated. When determining $\hat{p}_{i,C}$, she will therefore only rely on her noisy estimate $\hat{\alpha}_i$. When $\sigma_i = NR$, the agent chooses an incorrectly specified model of the world in which she believes ability is correlated with observable characteristics, and therefore $\hat{p}_{i,C}$ will be a function of both her noisy estimate $\hat{\alpha}_i$ and the social identity cue π_{θ_i}[13]. When the agent's subgroup is overrepresented among the successful individuals at the *Competitive* task in the society, *Not Repressing* the use of social identity in the decision-making process leads to an optimistic interpretation of the noisy estimate $\hat{\alpha}$, while this leads to a pessimistic interpretation of the noisy estimate when the agent's subgroup is underrepresented among the successful individuals in the society[14].

[13] See section 4 for a discussion on what happens when social identity has a direct effect on success.
[14] In the specific case where $\pi_\theta = \frac{1}{2}$, the two strategies *Repress* and *Not Repress* are equivalent.

I assume the estimate $\hat{p}_{i,C}$ is used in the decision-making process in the following way.

ASSUMPTION 2: *Let $\hat{p}^{\sigma_i}_{i,C}$ be the estimate of the probability of success of a Competitive task for agent i choosing σ_i. This agent chooses $a = C$ if and only if $\hat{p}^{\sigma}_{i,C} > \gamma$.*

This assumption can be interpreted in two ways. Either, agents are subjectively rational given the process that determines their subjective beliefs. Or, agents are boundedly rational in the sense that not all belief formation processes can be compared, meaning not all possible functions of $\hat{\alpha}$ and π_θ[15].

Assumption 2 also allows for two different interpretations of the model. One interpretation is that the choice of the model of the world alters the formation of the agents' beliefs about $\hat{p}_{i,C}$, where $\hat{p}_{i,C} \in \{\hat{p}^{R}_{i,C}, \hat{p}^{NR}_{i,C}\}$. I make nevertheless no assumptions about whether the agent actively forms beliefs and another interpretation of the model is that the use of social identity cues alters choice. Specifically, Assumption 2 is equivalent to saying that I consider a class of decision rules in which the agent specifies a threshold, such that she chooses the *Competitive* task if and only if her noisy estimate $\hat{\alpha}_i$ is above this threshold. The strategy set $\sigma_i \in \{R, NR\}$ implies that the agent can only choose between two thresholds. Consequently, an equivalent decision rule to the one presented in Assumption 2 is that an agent chooses $a = C$ if and only if $\hat{\alpha}_i > \gamma_i$, where

$$\gamma_i = \begin{cases} \gamma & \text{when } \sigma_i = R \\ \frac{\gamma}{\eta(\pi_{\theta_i} - p_{\theta_i})} & \text{when } \sigma_i = NR \end{cases} \tag{5}$$

The use of the social identity cue in the decision-making process implies therefore that the agent inflates or deflates the threshold for $\hat{\alpha}$ above which she thinks she is 'good enough' to undertake the competitive task. When following this interpretation of the model, the strategy set can also be directly specified as the choice set $\gamma_i \in \{\gamma, \frac{\gamma}{\eta(\pi_{\theta_i} - p_{\theta_i})}\}$, where the choice set is different for agents with different observable characteristics θ.[16]

[15]If agents could compare all such belief formation processes, they would behave as a Bayesian and choose $a = C$ when $\alpha > \gamma$. Because α and θ are independently distributed, a Bayesian analysis would be degenerate in this case. The model shows therefore the difference with a Bayesian model, by analysing whether, when agents are not able to draw all inferences given the structure of the model, this can open the door for agents to use information that is irrelevant, but that could still improve decision making.

[16]The fact that agents can only choose from two different thresholds can be considered as a modelling device that helps to keep the model parsimonious. One could add variability to α, to γ or to both and allow for a larger set of thresholds the agent can choose from, or equivalently, a larger family of belief formation processes based on $\hat{\alpha}$ and π_θ. A richer model would nevertheless lead to similar insights.

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

$$\Phi_{\alpha,\theta,\sigma_i,\Pi} = P(\hat{p}^{\sigma}_{i,C} > \gamma|\alpha) \tag{6}$$

Consequently, the expected pay-off for agent i of type $\{\alpha,\theta\}$ playing σ_i given a social context Π is,

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

with $\sigma_i \in \{R, NR\}$. The optimal strategies can now be defined as follows.

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

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

Definition 1 means that given the social context Π, an agent of type (α_i, θ_i) is assumed to use her social identity cue optimally on average over all possible realizations of $\hat{\alpha}_i$, where it follows from Equation (6) that the optimal strategy σ_i^* is adapted to α_i as if α_i were known. This assumption can be justified with a reinforcement learning process in which agents are able to recollect their outcomes and actions, but not the beliefs driving these actions. The true ability level α_i determines the outcomes the agent observes, which enables her to learn whether it is optimal to *Repress* or *Not Repress* the use of the social identity cue without precise knowledge of the relationship between her choice of strategy, choice of task and the observed outcome. Agents learn from their own experience and the optimal strategy σ_i^* defines in an implicit way what the agent knows about α_i. Because the set of strategies is small, this is easy for agents to calculate[17]. We shall see in Section 4 how the results extend to the case in which such learning would be imperfect.

[17]When agents choose $\sigma_i = NR$, they choose to believe in an incorrectly specified model of the world. I make no assumptions about whether agents know that the social identity cue contains no relevant information about their probability of success when undertaking a *Competitive* task. Therefore, if agents know the social identity cue is informationally irrelevant, the model can be interpreted as a model of optimal self-deception. On the other hand, when agents are unable to learn that the social cue is irrelevant and if, when agents identify as θ, they make themselves belief that the probability of success of agents like them has some predictive value for their own probability of success, the model could be interpreted as agents learning the optimal attribution error.

Population Equilibrium - The way in which agents use their social identity cues affects their choice behavior and therefore the social context Π. Let σ be the collection of σ_i. Each collection of strategies σ and social context Π generate choices and successes that in turn generate public data $\tilde{\Pi}$ such that,

$$\tilde{\pi}_\theta(\sigma, \Pi) = \frac{p_\theta \int \alpha \Phi_{\alpha,\theta,\sigma,\Pi} f(\alpha) d\alpha}{\sum_{\theta \in \Theta} p_\theta \int \alpha \Phi_{\alpha,\theta,\sigma,\Pi} f(\alpha) d\alpha} \tag{8}$$

where $f(\alpha)$ is the probability density function of α and $\tilde{\pi}_\theta(\sigma, \Pi)$ 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,*

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

In other words, a population equilibrium arises when all agents play their optimal strategy and this behavior induces a fixed point in the social context.

3 The Results

3.1 The Optimal Use of Social Identity

When an agent chooses the strategy *Repress*, this agent does not use social identity in decision making, while when the agent chooses the strategy *Not Repress*, this agent will use her social identity in the decision-making process. Definition 1 shows that the optimal use of social identity in this model is defined as the strategy that maximizes expected utility. I will illustrate this behavior with the following example.

Example - Consider a firm in which agents make a decision whether to pursue a career in management (competitive task) or to pursue a clerical job (outside option). Furthermore, assume that these agents can observe the current pool of successful managers and their gender, and that women are underrepresented in this pool. Let $\theta = 0$ denote being a woman and $\theta = 1$ being a man and assume $p_1 = p_0$. Let $\hat{p}^\sigma_{i,C}(\alpha, \theta)$ denote the estimate $\hat{p}_{i,C}$ implied by an agent of type $\{\alpha, \theta\}$ playing strategy σ. To illustrate behavior, consider only agents that have a true ability level $\alpha > \gamma$, where γ represents the probability of success for a clerical job. The welfare-maximizing choice for these agents is to pursue a management career and therefore, in any learning process, agents will observe more successful outcomes when choosing the strategy that makes them most likely to choose this career. When agents decide to *Repress* the urge to look at others like them when forming their estimate of becoming a successful manager, the probability with which they choose to pursue a management career is equal to $P(\hat{\alpha} > \gamma)$. It is therefore only optimal to *Not Repress* the urge to look at others when $P(\hat{p}_{i,C}^{NR} > \gamma) \geq P(\hat{\alpha} > \gamma)$.

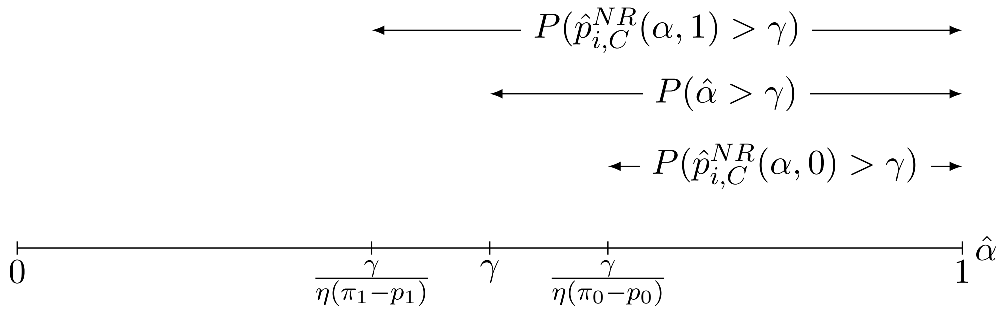

Figure 1: An Illustration of the different probabilities to choose $a = C$ for an agent with $\alpha > \gamma$ in a social context such that $\pi_0 < \pi_1$.

Figure (1) illustrates the implications of the different strategies *Repress* and *Not Repress* for both genders. Because men are overrepresented among the current successful managers in the firm, using the social identity cue in decision making causes men to deflate the threshold above which they think they are 'good enough' to become a successful manager. In other words, choosing a management career becomes relatively more attractive than choosing a clerical job. Men with $\alpha > \gamma$ will therefore learn it is optimal to *Not Repress* the urge to look at others. For women the story is different. Because women are underrepresented among the successful managers, using the social identity cue in the formation of an estimate of their likelihood of success as a manager would inflate the threshold above which they think they are 'good enough'. This would make a management career relatively less attractive and, in order to maximize their expected utility, women will therefore learn to *Repress* the urge to look at the outcomes of other women. The opposite reasoning applies for agents with a true ability level $\alpha < \gamma$. In general, one can show that agents with a relatively high ability type will choose to take into account their social identity when they belong to a socially more successful group, while they will wish to avoid it when they belong to a socially less successful group, and vice versa for agents with a relatively low ability type. Proposition 1 formalizes this result.

PROPOSITION 1 (The Optimal Use of Social Identity): *The optimal strategies σ^* given an agent's type $\{\alpha, \theta\}$ are the following:*

- *For agent's of type $\{\alpha, \theta\}$ such that $\alpha > \gamma$ and $\pi_\theta > p_\theta$ or $\alpha < \gamma$ and $\pi_\theta < p_\theta$, the optimal strategy σ^* is 'Not Repress'*
- *For agent's of type $\{\alpha, \theta\}$ such that $\alpha > \gamma$ and $\pi_\theta < p_\theta$ or $\alpha < \gamma$ and $\pi_\theta > p_\theta$, the optimal strategy σ^* is 'Repress'*

Proposition 1 demonstrates that agents use their social identity to improve decision making on average by nudging their beliefs in the correct direction and therefore shows that, when agents are limited in their ability to make inference, it becomes useful to incorporate signals in the decision function that would otherwise be ignored. This shows how choice behavior can be driven by observable characteristics, even though these observable characteristics have no direct effect on utility.

Talent will always find its way - The ability to improve decision making using social identity is a function of the true ability level α. Specifically, the model predicts that the behavior of agents with ability levels close to γ will be influenced much more by the social context than the behavior of agents with extreme ability levels. To illustrate this, consider an agent with $\alpha > \gamma$. The probability for such an agent to make the incorrect decision when not using social identity is given by $P(\hat{\alpha} < \gamma)$. Because $\hat{\alpha}$ is unbiased, the further away the true ability level α is from γ, the lower the probability that the agent receives a noisy estimate $\hat{\alpha} < \gamma$[18]. A similar reasoning applies to agents with $\alpha < \gamma$[19]. Therefore, the use of social identity is most beneficial for those who have an ability level close to γ, while agents with extremely low or extremely high ability levels are always more likely to take the correct choice, independent of their observable characteristics and the social context in which they make their decisions.

Confidence Management - The intermediate mechanism through which the use of social identity affects decision making is the optimistic or pessimistic processing of the noisy estimate $\hat{\alpha}$, where agents process their noisy estimate differently depending on their observable characteristic. This optimism or pessimism can enhance or offset the agent's over- or under-confidence regarding her chances of success that is captured in

[18] The larger the difference $(\alpha - \gamma)$, the more the area below γ moves towards the tail of the distribution and the lower the probability $G_\alpha(\gamma)$.

[19] The larger the difference $(\gamma - \alpha)$, the lower the probability $1 - G_\alpha(\gamma)$.

the noisy estimate $\hat{\alpha}$. The use of social identity is therefore a mechanism that gives the agent the option to manage confidence and Proposition 1 shows that the optimal use of social identity results in the optimal management of confidence to improve decision making. The model therefore sheds light on the instrumental value of both over- and under-confidence to improve decision making[20]. Because the option to manage confidence using the social identity cue depends on the observable characteristic of the agent, the option to improve decision making using the social identity cue is not available to all types of agents and the same agent can have different options in different social contexts.

3.2 The Social Context in Equilibrium

The asymmetry in the available options to manage confidence across different social types could potentially create asymmetries in choice behavior. In this section, we analyze the effects the optimal use of social identity cues on choice behavior at the aggregate level.

Selection and Population Effects - The optimal use of social identity in a social context in which one subgroup is overrepresented among the successful individuals affects both how many and what type of agents choose a *Competitive* task. Specifically, if an observable characteristic implies a more pessimistic processing of the noisy estimate, those who choose the *Competitive* task despite this, tend to have a larger success rate on average than those who choose *Competitive* task with an observable characteristic that implies optimistic processing of the noisy estimate. This is what we call the '*selection effect*'. On the other hand, the population of those that belong to the socially less successful subgroup and that choose the *Competitive* task tends to be smaller than the population of those that choose the *Competitive* task and belong to the socially more successful subgroup.

[20]At the same time, the model does not contradict the fact that agents also derive a form of hedonic value from self-esteem through their self-image. This is trivial for agents who use optimistic processing that leads to over-confidence (Mobius et al., 2014). For agents that use pessimistic processing that leads to under-confidence, this can be explained using self-affirmation theory (Sherman and Cohen, 2006). Self-affirmation theory states that agents have a sort of psychological immune system that protects their ego or self-image from information that could lead to a negative evaluation of oneself. This psychological immune system creates all kinds of reactions to possible threats to the self that make it possible for the individual to process the information in a way that will not hurt their self image. In this model, by identifying as a group member, agents attribute the fact that they are not good enough to undertake the task to group characteristics, instead of to individual-specific factors. Through this type of causal attribution, agents are able to acknowledge that they are 'not good enough' to undertake the competitive task, while still keeping their self-image in tact.

COROLLARY 1: *Let $\theta' \in \Theta$ be the complement of θ and assume WLOG that $\pi_\theta > \pi_{\theta'}$. The optimal use of social identity has both a population effect, such that $\Phi_{\alpha,\theta,\sigma_i,\Pi} > \Phi_{\alpha,\theta',\sigma_i,\Pi}$ and a selection effect, such that $E(\alpha|\hat{p}^{\sigma_i} > \gamma, \theta) < E(\alpha|\hat{p}^{\sigma_i} > \gamma, \theta')$. The strength of both effects is such that the order $\pi_\theta > \pi_{\theta'}$ will always be preserved.*

Example - To illustrate the selection and population effect, consider again the firm in which agents have to choose to pursue a career in management or a clerical job, and the relevant observable characteristic is gender. As before, we are in a social context where women are underrepresented among the successful managers, and where $\theta = 1$ represents being a men and $\theta = 0$ represents being a women. Proposition 1 implies that men with $\alpha > \gamma$ will use their social identity, while women with $\alpha > \gamma$ will not. Similarly, women with $\alpha < \gamma$ will use their social identity, while men with $\alpha < \gamma$ will not.

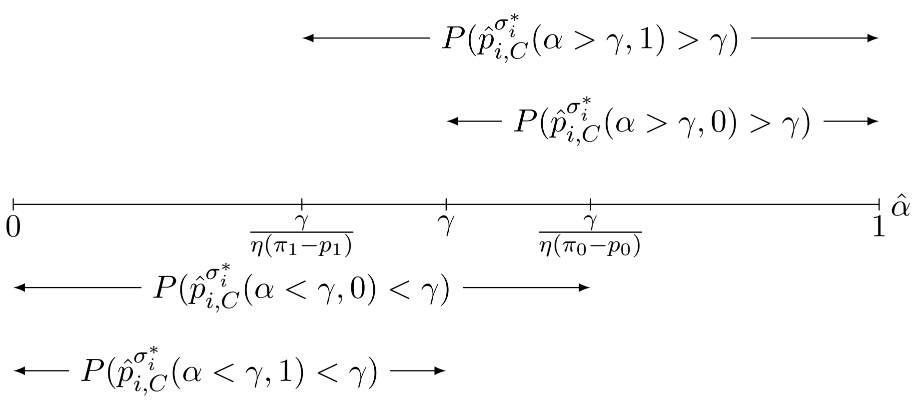

Figure 2: Optimal Decision-Making in a social context such that $\pi_0 < \pi_1$. The part above the axis describes for which realizations of $\hat{\alpha}$ agents correctly choose $a = C$, while the part under the axis shows for which realizations of $\hat{\alpha}$ agents correctly choose $a = NC$.

Figure (2) shows the probabilities with which men and women will choose to pursue a career in management or a clerical job. Because the number of agents in the society is arbitrarily large, these probabilities can be interpreted as population fractions. Women with an ability level $\alpha > \gamma$ will only choose to pursue a career in management when $\hat{\alpha} > \gamma$, while men with an ability level $\alpha > \gamma$ will choose to do so when $\hat{\alpha} > \frac{\gamma}{\eta(\pi_1 - p_1)}$. Therefore, more men with $\alpha > \gamma$ will choose to pursue a career in management than women. Similarly, more women with $\alpha < \gamma$ will choose a clerical job than men with $\alpha < \gamma$. Furthermore, men with $\alpha < \gamma$ will mistakenly choose a career in management when they receive a realization $\hat{\alpha} > \gamma$, while women with $\alpha < \gamma$ will only choose to do so when they receive a realization $\hat{\alpha} > \frac{\gamma}{\eta(\pi_0 - p_0)}$. Since $\gamma < \frac{\gamma}{\eta(\pi_0 - p_0)}$, the latter probability is smaller. Similarly, women with $\alpha > \gamma$ will mistakenly choose a clerical job when they receive a realization $\hat{\alpha} < \gamma$, while men with $\alpha > \gamma$ will only choose a clerical job when they receive a realization $\hat{\alpha} < \frac{\gamma}{\eta(\pi_1 - p_1)}$. Again, Figure (2) shows that the latter event is less likely. We therefore conclude that more men will, both correctly and mistakenly, choose to pursue a management career, while more women will, both correctly and mistakenly, choose a clerical job. This demonstrates the population effect. Finally, higher realizations of $\hat{\alpha}$ are more likely for agents with higher ability levels α. Because $\gamma > \frac{\gamma}{\eta(\pi_1 - p_1)}$, men choose the management task for on average lower realizations of $\hat{\alpha}$ than women and therefore, $E(\alpha | \hat{p}^{\sigma_i} > \gamma, 1) < E(\alpha | \hat{p}^{\sigma_i} > \gamma, 0)$. In other words, the average success rate of men in a management career will be lower than the average success rate of women, which demonstrates the selection effect [21]. The intuition behind the result that the selection and population effect will not reverse the order $\pi_1 > \pi_0$, is that, because α and θ are independently distributed, the norm of the set of agents of type $\{\alpha, 1\}$, such that $\hat{\alpha} > \gamma$ is equal to the norm of the set of agents of type $\{\alpha, 0\}$, such that $\hat{\alpha} > \gamma$. The population of men that choose a management career consists therefore of the men that observe $\hat{\alpha} > \gamma$ plus the men that observe $\frac{\gamma}{\eta(\pi_1 - p_1)} < \hat{\alpha} < \gamma$, while the population of women that chooses a management career only consists of those women that observe $\hat{\alpha} > \gamma$.

Population Equilibria - What we could observe in equilibrium depends on whether the population and selection effect make the difference $\pi_\theta - \pi_{\theta'}$ shrink or increase. There are two foreseeable scenarios that could appear in equilibrium. On the one hand, there could be a population equilibrium in which social identity does not drive choice behavior. Since α and θ are independently distributed, such a population equilibrium implies a symmetric allocation of agents belonging to different subgroups over tasks. On the other hand, the selection and population effect that are induced by the social context could induce asymmetric allocations over tasks across a priori identical subgroups. These two different scenarios are defined in Definition 3.

[21] The selection effect is driven by the earlier described result that observable characteristics especially drive the choice behavior of people with true ability levels close to γ. Agents with extreme ability levels are more likely to make the correct choice independent of their observable characteristic.

DEFINITION 3: *A population equilibrium in which the use of social identity in decision making is the same and the allocation of individuals over the competitive and non-competitive task is symmetric across different subgroups is called a 'Neutral Regime'. A population equilibrium in which the use of social identity in decision making is different and the allocation of individuals over the competitive and non-competitive task is asymmetric across different subgroups is called a 'Non-Neutral Regime'.*

In the following, I analyze under what conditions a 'Neutral Regime' and a 'Non-Neutral Regime' can exist. Because α and θ are independently distributed, it is easy to see that a population equilibrium with a 'Neutral Regime' always exists. To illustrate how a 'Non-Neutral Regime' can arise, consider the following extreme case.

Example - Consider again the example of a firm with the same share of male and female agents who choose between pursuing a management career and a clerical job. Now, also assume agents have the following extreme reaction function,

$$\eta(\pi_\theta - p_\theta) = \begin{cases} +\infty & \text{if } \pi_\theta > p_\theta \\ 1 & \text{if } \pi_\theta = p_\theta \\ -\infty & \text{if } \pi_\theta < p_\theta \end{cases} \tag{10}$$

When $\pi_0 = \pi_1 = \frac{1}{2}$, the strategies *Repress* and *Not Repress* are equivalent. There will therefore be no asymmetry in the choice behavior of men and women and this social context induces a social context such that $\tilde{\pi}_1(\pi_1 = \frac{1}{2}, \sigma^*) = \frac{1}{2}$. In other words, a 'Neutral Regime' exists. Nevertheless as soon as agents observe slightly more men than women among the successful managers, such that $\pi_1 > \pi_0$, the extreme reaction function $\eta(\pi_\theta - p_\theta)$ will make all men with $\alpha > \gamma$ choose to pursue a management career, while all women with $\alpha < \gamma$ will choose the clerical job. Consequently, $\tilde{\pi}_1(\pi_1, \sigma^*) > \pi_1$, while $\tilde{\pi}_0(\pi_0, \sigma^*) < \pi_0$ and the 'Neutral Regime' becomes unstable. In this extreme case, the social identity cue π_1 converges to its upper bound. Let S_1 be the number of successful male managers. The upper bound $\overline{S}_1$ on S_1 is characterized by the fact that, with this extreme reaction function, all men with $\alpha > \gamma$ and all men with $\alpha < \gamma$, but $\hat{\alpha} > \gamma$ choose to pursue a management career, and is equal to,

$$\overline{S}_1 = \int_{\alpha<\gamma} \int_{\hat{\alpha}>\gamma} \alpha g_\alpha(\hat{\alpha}) f(\alpha) d\hat{\alpha} d\alpha + \int_{\alpha>\gamma} \alpha f(\alpha) d\alpha \tag{11}$$

Similarly, let S_0 be the number of successful female managers. The lower bound $\underline{S}_0$ on

S_0 is characterized by the fact that, with this extreme reaction function, only women with $\alpha > \gamma$ and $\hat{\alpha} > \gamma$ and no women with $\alpha < \gamma$ will choose to pursue a management career. Consequently,

$$\underline{S}_0 = \int_{\alpha>\gamma} \int_{\hat{\alpha}>\gamma} \alpha g_\alpha(\hat{\alpha}) f(\alpha) d\hat{\alpha} d\alpha \tag{12}$$

The upper bound on the induced social context $\tilde{\pi}_1(\pi_1, \sigma)$ is now given by

$$\overline{\pi}_1(\pi_1, \sigma) = \frac{\overline{S}_1}{\overline{S}_1 + \underline{S}_0} \tag{13}$$

Because $\underline{S}_0 > 0$, it follows that $\overline{\pi}_1(\pi_1, \sigma)$ is always strictly below 1. This illustrates that, even with an extreme reaction function, in a 'Non-Neutral Regime' the pool of successful managers will always consist of both female and male agents.

Proposition 2 shows under what condition the 'Neutral Regime' becomes unstable when you disturb it locally for more general specifications of the reaction function $\eta(\pi_\theta - p_\theta)$.

PROPOSITION 2: *Assume $p_0 = p_1$. A population equilibrium with a 'Neutral Regime' always exists and can co-exist with a stable population equilibrium with a 'Non-Neutral Regime' when the following condition is satisfied,*

$$\gamma > \frac{2}{\epsilon_{S,\gamma}(0) \frac{\partial \eta(0)}{\partial \pi_\theta}} \tag{14}$$

where S is the total number of individuals that have successfully completed the Competitive task and $\epsilon_{S,\gamma}(0) = -\frac{\frac{\partial S}{\partial \gamma}|_{\pi_\theta = p_\theta}}{S}$.[22]

For more general specifications of the reaction function, Proposition 2 shows that there are three different ingredients that can make a 'Neutral Regime' unstable. First of all, γ has to be sufficiently large. This means that, for a 'Non-Neutral Regime' to exist, the *Non-Competitive* task must be sufficiently attractive relative to the *Competitive* task. The intuition behind this is that, when γ is large, only few agents will undertake and manage to be successful at the *Competitive* task and the effect of *Not Repressing* the use of π_θ on the decision making of agents can have a strong impact on the induced social identity cue $\tilde{\pi}_\theta$. Secondly, a 'Neutral Regime' becomes unstable when the steepness of

[22]Because $\frac{\partial S}{\partial \gamma}|_{\pi_\theta = p_\theta} < 0$, the value $\epsilon_{S,\gamma}(0)$ is always strictly positive.

the reaction function $\frac{\partial \eta(0)}{\partial \pi_\theta}$ is sufficiently large. When this is the case, an agent's belief formation process when choosing to *Not Repress* the use of π_θ in decision making will react strongly enough to even small deviations of π_θ from p_θ. A last ingredient for a 'Neutral Regime' to become unstable is the elasticity $\epsilon_{S,\gamma}(0)$ of the total number of successful individuals at the *Competitive* task S with respect to γ evaluated in this 'Neutral Regime'. This elasticity has to be sufficiently large, which is the case when S is small. Therefore this elasticity is increasing in γ. Furthermore, the elasticity depends positively on $g_\alpha(\gamma)$, which we can write as

$$g_\alpha(\gamma) = lim_{t\to 0}\frac{1}{2t}P(|\hat{\alpha} - \gamma| < t|\alpha) \tag{15}$$

and can be interpreted as the 'relative likelihood' that people observe a realization of $\hat{\alpha}$ very close to γ given their true ability level α. When this 'relative likelihood' is high, the optimal strategies of agents can have instrumental value in decision making even for small deviations of π_θ from p_θ. Finally, there is a complementarity between γ and $\frac{\partial \eta(0)}{\partial \pi_\theta}$, such that, for higher values of γ, we need lower values of $\frac{\partial \eta(0)}{\partial \pi_\theta}$ for the sufficient condition to be satisfied and vice versa.

In Proposition 2, we derive the sufficient condition for the existence of a 'Non-Neutral Regime' when the fractions of the population having an observable characteristic θ are the same. We now analyze what happens when this is not the case.

Example - Consider again the firm of the previous examples, but assume now that there are more male than female agents. Because α is distributed over the population independently of gender, the fact that there are more men than women in the firm also implies that there will be more men than women with an ability level $\alpha > \gamma$. Also, assume again that we are in a social context, such that men are overrepresented among the pool of successful managers. Because $p_1 > p_0$, this means that $\pi_1 > p_1$, while $\pi_0 < p_0$. Proposition 1 shows that, in such a social context, all men with $\alpha > \gamma$ will choose to *Not Repress* the use of the fact that men are overrepresented among the successful managers in their decision-making process, while all women with $\alpha > \gamma$ will choose to *Repress* the use of the fact that women are underrepresented among the successful managers. Because $p_1 > p_0$, the fraction of the population that can use the social identity cue to increase the likelihood they will choose the *Competitive* task correctly is now larger than in the case when $p_1 = p_0$. Therefore, when there are more men than women in the firm, a social context in which men are overrepresented among the successful managers induces a larger fraction of successful male managers than when the population of male and female agents in the firm is equal. This result is formalized in Corollary 2.

COROLLARY 2: *The set of values* γ *and reaction functions* η *for which the sufficient condition of Proposition 2 for the existence of a population equilibrium with a 'Non-Neutral Regime' in which* $\pi_1 > p_0$ *is satisfied increases when* $p_1 > p_0$.

Furthermore, as presented in the example with the extreme reaction function η, the social identity cue that can exist in equilibrium is bounded from above. Corollary 3 sheds more light on the relation between the reaction function η and the degree of asymmetry in a 'Non-Neutral Regime', given that this regime exists.

COROLLARY 3: *Take two reaction functions* $\hat{\eta}$ *and* η*, such that* $\hat{\eta}(\pi_\theta - p_\theta) > \eta(\pi_\theta - p_\theta)$ *for all* $\pi_\theta > p_\theta$*. Assume WLOG that a 'Non-Neutral Regime' exists in which* $\pi_1 > p_1$*. Let* $\pi^*_{\eta,1}$ *be the equilibrium value of* π_1 *given a reaction function* η*. Then,* $\pi^*_{\hat{\eta},1} > \pi^*_{\eta,1}$.

Finally, without any extra assumptions on $\eta(\pi_\theta - p_\theta)$, a population equilibrium with a 'Neutral Regime' is not stable whenever it co-exists with a population equilibrium with a 'Non-Neutral Regime'. If we want both population equilibria to be stable when they co-exist, we need to assume that agents are not able to notice differences between π_θ and p_θ unless they are striking, meaning that the function $\eta(\pi_\theta - p_\theta)$ needs to be flat when π_θ is ϵ-close to p_θ.[23].

While Proposition 1 shows that the optimal use of social identity is an optimal behavior from an individual perspective, Proposition 2 shows that this is not necessarily the case from the perspective of the society as a whole. The option to optimally bias decision making using the social identity cue is different for different social types depending on the social context, and at the aggregate level this can cause a different allocation over tasks across a priori identical subgroups. Especially for choice contexts in which only a small part of the population is successful at the *Competitive* task and in which the outside option is relatively attractive, the optimal use of social identity can make a 'Neutral Regime' unstable. This causes the existence of a 'Non-Neutral Regime' that en-

[23]This can be obtained by introducing a function

$$\eta^0(\pi_\theta - p_\theta) = \begin{cases} 1 & \text{if } |\pi_\theta - p_\theta| < \epsilon \\ \eta(\pi_\theta - p_\theta) & \text{otherwise} \end{cases} \tag{16}$$

ables any asymmetry in the representation of individuals belonging to a priori identical subgroups among the successful individuals at the *Competitive* task to persist in equilibrium. Moreover, a 'Non-Neutral Regime' in which a certain subgroup is overrepresented among the successful individuals exists for a larger set of values γ and reaction functions η when this subgroup represents a majority of the population, and the higher the curve of the reaction function η, the more asymmetry we will observe in equilibrium. Nevertheless, even in the most extreme case, the equilibrium pool of successful individuals at the *Competitive* task will always consist of agents belonging to both subgroups.

4 Discussion

4.1 A Misspecified Reaction Function

In the exposition of the model, I assume that agents have a correctly specified reaction function $\eta(\pi_\theta - p_\theta)$. This assumption requires nevertheless that agents hold correct beliefs about the fraction p_θ of the population that has an observable characteristic θ. It is not always obvious that this is the case.

Example - Consider again the firm in which male and female workers have to make a choice between pursuing a career in management or a clerical job. Assume that there are less women than men that have the qualifications to pursue a management career, but that agents fail to take this into account. In other words, people hold a belief $\hat{p}_0 > p_0$ about the fraction of qualified women in the population. Specifically, let $\hat{p}_0 = \frac{1}{2}$, while $p_0 < \frac{1}{2}$. In this case, agents expect to observe the same fraction of men and women in the pool of successful managers, while in reality this is not what a 'Neutral Regime' looks like. This means that, when a 'Neutral Regime' appears, agents will not interpret it as such and they will perceive women to be underrepresented, while men are perceived to be overrepresented. This will induce the differential use of social identity cues, which will drive the population automatically away from the 'Neutral Regime'. This result is formalized in Corollary 4.

COROLLARY 4: *Assume WLOG that agents hold an incorrect belief $\hat{p}_\theta > p_\theta$. In this case, a population equillibrium with a 'Neutral Regime' no longer exists and there will only exist a population equilibrium with a 'Non-Neutral Regime' in which $\pi_\theta < p_\theta$.*

Corollary 4 also shows that the only population equilibrium that can exist in the abovementioned example, is a population equilibrium with a 'Non-Neutral Regime' in which women are indeed underrepresented among the successful managers. A misspecified reaction function has therefore a significant effect on the possible equilibrium outcomes and makes a misspecified interpretation of the social context self-fulfilling.

4.2 Imperfect Learning

In Definition 1, I assume that agents are perfectly able to learn the optimal strategies that maximize their expected utility. The main point of this assumption is to show that, even when agents are able to perfectly learn how to optimally use the social identity cues, asymmetries can still persist in equilibrium. The assumption can nevertheless be easily challenged. In this section, I discuss that, as long as the probability to learn is positive, imperfections in the learning process will not necessarily invalidate the result of the existence of a population equilibrium with a 'Non-Neutral Regime'.

As mentioned in the exposition of the model, the behavior of agents can be justified with a reinforcement learning process in which agents are able to learn their optimal strategies without knowing their true ability level α through learning from their own experience. In such a learning process, agents are only able to recollect their outcomes and actions, but not the beliefs that drove these actions. The true ability level α will determine what agents observe in the learning process. The equilibrium model can be adjusted to allow for imperfect learning as follows. The induced probability to choose the competitive task for an agent of type $\{\alpha, \theta\}$ in a social context Π, playing strategy σ_i as presented in Equation (4) can be written as,

$$\Phi_{\alpha,\theta,\sigma_i,\Pi} = \Sigma_{\sigma \in \{R,NR\}} P(\sigma_i = \sigma | \alpha, \theta, \Pi) P(\hat{p}^{\sigma}_{i,C} > \gamma | \alpha) \tag{17}$$

where, in the case of perfect learning, $P(\sigma_i = \sigma | \alpha, \theta, \Pi) \in \{0, 1\}$, while in the case of imperfect learning, $P(\sigma_i = \sigma | \alpha, \theta, \Pi) \in [0, 1]$. Let us now specify an exogenous learning process λ. Then, any such learning process implies a probability $P^{\lambda}(\sigma_i = \sigma | \alpha, \theta, \Pi)$. For example, let $v(\sigma)$ be the expected pay-off of playing strategy σ and let $\lambda = QR$ represent a quantile response model. Then, $P^{QR}(\sigma_i = \sigma | \alpha, \theta, \Pi) = \frac{exp(\beta v(\sigma))}{\Sigma_{\sigma \in \{R,NR\}} exp(\beta v(\sigma))}$, and the larger β, the better the agent is able to learn. Similarly, we can assume $\lambda = S$ represents a sampling model in the spirit of Osborne and Rubinstein (1998). In that case, $P^{S}(\sigma_i = \sigma | \alpha, \theta, \Pi) = P(\hat{v}(\sigma) > \hat{v}(\sigma'))$, where $\hat{v}(\sigma) = v(\sigma) + \epsilon$ represents the noisy value that agents observe through sampling. The larger the number N of samples the agent takes into account, the smaller the value of ϵ and the better the agent is able to learn. In both cases, the probability to learn in strictly positive.

Now, as long as agents are able to learn something, there will be a discrepancy between $\Phi_{\alpha,\theta,\sigma_i,\Pi}$ for different values of θ, which will automatically induce differences in choice behavior. Furthermore, a strategy σ_i only has instrumental value when playing $\sigma_i = R$ implies a choice of action $a_i \in \{C, NC\}$ that is different from the choice implied by playing $\sigma_i = NR$. Therefore, only in these instances agents will be able to learn and these are more likely to happen to agents with ability levels close to γ. As discussed earlier, these agents also happen to be the agents who can potentially gain most from the use of social identity. Therefore, the agents through whom the optimal use of social identity has the largest effect at the aggregate level, will also be the agents that learn best in a case of imperfect learning.

Finally, to discuss the effect of imperfect learning on the possible equilibrium outcomes, we also need to discuss how we assume agents learn. In the exposition of the model, I assume that agents have a natural 'urge' to look at others when they are not sure what to do or believe. If failing to learn means failing to *Repress* the use of social identity when this is optimal, imperfect learning implies that more agents than optimal use their social identity cue. This would increase the set of values of γ and functions $\eta(\pi_\theta - p_\theta)$ for which a population equilibrium with a 'Neutral Regime' and a 'Non-Neutral Regime' can co-exist. If failing to learn implies that agents make random mistakes, the set of values of γ and functions $\eta(\pi_\theta - p_\theta)$ for which a population equilibrium with a 'Neutral Regime' and a 'Non-Neutral Regime' can co-exist will be smaller than in the case of perfect learning.

4.3 Correlated Ability and Observable Characteristics

To isolate the effect of the optimal use of social identity on decision making, I assume the ability type and the observable characteristic are independently distributed over the population. In practice, a high ability type often results from the degree to which this ability has been developed by the agent. The economic literature provides plenty of empirical evidence for a correlation between observable characteristics and the available opportunities to develop certain abilities, that leads to correlated (developed) ability types and observable characteristics[24]. Such a correlation will not affect the potential gains of using social identity in decision making at the individual level, and will therefore not affect the optimal strategies as presented in Proposition 1. It will nevertheless affect

[24]See for example Carneiro et al. (2003), who provide evidence of differences in premarket skills across different race groups.

the outcomes observed at the aggregate level. In most empirical examples, having a high ability type is correlated with belonging to the socially more successful subgroup in the social context. In this model, such a correlation increases the total number of agents that can use their social identity to improve decision making. Therefore, the use of social identity would reinforce any differences in the prevalence of the different subgroups among the successful individuals that are caused by an unequal distribution of skills over different social types. The effects of the optimal use of social identity on the social context will therefore complement the effects of this type of correlation between ability and observable characteristics. It could also be that the correlation works in the opposite direction, where having a low ability type is correlated with belonging to the socially more successful subgroup. In this case, the number of agents that will be able to use social identity to improve decision making decreases. Moreover, at the aggregate level, the effect of the optimal use of social identity and the effect of correlated ability and observable characteristics work in opposite directions. This type of correlation will therefore decrease the set of values of γ and functions η for which a population equilibrium with a 'Non-Neutral Regime' exists.

4.4 Correlated Probability of Success and Observable Characteristics

Although not considered in this paper, social identity also has direct effects on utility. Agents can derive for example utility from the acceptance by her peers (Austen-Smith and Fryer, 2005) or they can lose utility due to the punishment of other agents for not complying with social norms (Akerlof and Kranton, 2000). Furthermore, agents' beliefs can directly affect performance (Compte and Postlewaite, 2004) or performance can be affected by the beliefs of others due to discrimination (Coate and Loury, 1993). There could also be differences in the probability of success of different subgroups due to network effects (Lalanne and Seabright, 2011). This implies that the probability of success $p_{i,C}$ is not only a function of ability α, but also of the observable characteristic θ. The social identity cue now contains real information and the agent can use her social context to learn about her probability of success for a '*Competitive*' task. When the agent chooses to take the social identity cue into account, she chooses for a correctly specified model of the world and her belief formation process represents a process of social learning. Furthermore, when using the model of Compte and Postlewaite (2004), the optimal confidence management following from the use of social identity will affect the true probability of success in a way that will reinforce the asymmetries caused by the population and selection effect. Therefore, in this model, any direct effects of social identity on utility will be complementary to the effects of the optimal use of social identity, and both the direct effects of social identity on utility and the optimal use of social identity will be simultaneously and endogenously determined through the social context Π. The results presented in Propositions 1 and 2 are therefore robust in a context where social identity has a direct effect on utility and the set of values for which a population equilibrium with a 'Non-Neutral Regime' exists will increase.

4.5 Correlation Noise in Estimate Ability and Ability itself

I assume that the noise in the unmodelled inference process that is used to learn about ability is not correlated with the ability type itself. It is nevertheless likely that the inference process of agents with low ability types is more noisy than the inference process of agents with high ability types, because agents with high ability types are more likely to undertake tasks related to this ability and will therefore receive more feedback on their ability type. Moreover, the noise could also be correlated with the observable characteristic, because agents with an observable characteristics that makes them less likely to choose a *Competitive* task given their ability level are also less likely to receive feedback on their ability type. The model shows that agents that have a more noisy inference process can potentially gain more from the optimal use of social identity. For those agents that can use their social identity to improve decision making, a higher degree of noisiness in their inference process can therefore be partially off-set with the optimal use of social identity. Those agents that cannot use their social identity to improve decision making will not have an instrument to off-set their higher likelihood to make mistakes. At the aggregate level, the effects of these mistakes will nevertheless be complementary to the effects of the optimal use of social identity and will increase the set of values for which a population equilibrium with a 'Non-Neutral Regime' exists. Finally, one can imagine that agents with extreme ability levels learn through a less noisy inference process than agents with average ability levels. Because the effect of the optimal use of social identity affects the social context mainly through agents with average ability levels, this type of correlation will not affect the equilibrium results much.

5 Applications

5.1 Perceptions and Role Models

The model predicts that, when people are uncertain about their ability type, their choice behavior is driven by what they perceive in their social context. Moreover, the perception agents have of the social context can become self-fulfilling in equilibrium. The model therefore supports the belief that it is important to shed light on more women and minority people that are successful in areas in which these groups are underrepresented, which makes the paper related to the literature on role models. The effect of role models on decision making works nevertheless through a different mechanism than in Chung (2000), where agents use role models to retrieve information about the degree of hostility of the environment they will work in. In this model, agents use the social context, and therefore role models, to retrieve information regarding their own ability type, and the effect of role models on decision making works through the optimal management of confidence.

This result has interesting policy implications. It implies that people can be nudged towards desirable behavior by influencing how people perceive their social context. This could be achieved by influencing the structure on the publicly available data and statistics that people take into account, or by making hidden data more visible, and could be a policy complementary to a real and maybe more costly change of the social context, through for example affirmative action policy.

5.2 Individual Feedback

The use of social identity cues provides an option to optimally manage confidence, but this option is not available to all types of agents. The paper highlights therefore that similar types of individual feedback can have very different effects on future choice behavior for agents belonging to different social groups, and one could exploit these differences to create more diversity in educational and professional environments. The asymmetry in options to manage confidence using the social context could be partially offset by creating a similar bias in the noisy estimate of ability for exactly those agents that cannot optimally bias their decision making use the social identity cues. For example, if men are overrepresented among the successful managers, only men have the option to bias their estimate of success in a management career upwards using the social context. If one wants to induce more women to choose a career in management, this could be achieved by giving those women, that have the capabilities to become good managers, systematically more positive feedback regarding their abilities than men. This would bias their individual-specific noisy estimate of ability, and therefore their confidence, upwards in a similar way as what men can achieve with the use of the social context.

5.3 Stereotypes

Bordalo et al. (2016) define stereotypes as the subgroups that are most representative for a certain task. This model shows how stereotypes can be determined endogenously through the interaction between the choice behavior of agents and the social context in which they make these choices. The driving forces behind this process are the differential use of social identity cues and the ambiguity in the learning process of certain ability types. The asymmetries in choice behavior across different social groups, induced by the differential use of social identity cues, can be reinforced by discrimination, social pressure and other direct effects of social identity on utility, and provide the public data from which stereotypes will be derived. Consequently, the model has a similar prediction to Hoff and Stiglitz (2016) and Steele (2010), namely that stereotypes cannot survive in a society, unless they are a reflection of what people observe in their social context, and argues that the only way to effectively fight harmful stereotypes is to change the reality in which people make decisions. This is empirically demonstrated by Banaji and Greenwald (2016), who show that, now the workplace is populated by as many women as men, the implicit male-career association is growing ever fainter among the young. Gender stereotypes remain nevertheless present and they argue this is because of the strong and dominating position of women in the home sphere and men in the highest status positions at work.

5.4 Affirmative Action

The model sheds light on the necessary conditions for temporary affirmative action to have long-run effects. Specifically, the model provides a theoretical interpretation of the phenomenon 'critical mass'. This term refers to the point at which there are enough successful minorities in a setting, such that individual minorities no longer feel an interferencing level of identity threat (Steele, 2010). In this model, a 'critical mass' is defined as the quantity of successful minorities, for which minorities no longer perceive themselves as underrepresented among the successful individuals. This quantity is determined by the reaction function of agents, and specifically by their sensitivity to changes in the social context and their beliefs p_θ regarding what a 'Neutral Regime' should look like. If temporary affirmative action increases the representation of minorities among the successful individuals, but minorities still consider themselves being underrepresented, then, as soon as the policy is removed, choice behavior driven by the differential use of social identity cues will automatically bring the society back to a 'Non-Neutral Regime'. Temporary affirmative action will therefore not have any long-run benefits.

5.5 Oppositional Identities

The phenomenon of oppositional identities tries to describe why, in some minority groups, agents tend to reject the dominant culture. This type of behavior is for example observed in schools, and modelled using social pressure (Austen-Smith and Fryer, 2005) or cultural transmission (Bisin et al., 2011). This model proposes an alternative view on the phenomenon. Specifically, within this framework, an oppositional identity can be defined as follows. Assume the '*Competitive*' task represents a task that is the social norm according to the dominant culture in society, for example working hard in school, while the '*Non-Competitive*' task is something that is frowned upon, for example shirking in school. Following the literature on oppositional identities, we would say that students that choose to shirk in school choose an oppositional identity. This choice would become a trait of an entire minority group, when a disproportionate amount of students belonging to the minority group chooses this option. The model shows that, even in the absence of social pressures or cultural transmission, we would observe this type of choice behavior among minority students when these students observe that students like them are underrepresented among the successful students that work hard in school. The model argues therefore that the observation, that minorities disproportionately often adopt an oppositional identity, can also be the result of what these minorities observe in their social context. Specifically, choosing an oppositional identity, when it is not optimal for one to do so, is driven by exhibiting a lack of confidence regarding one's own ability and not having the option to offset this lack of confidence using information from the social context.

6 Conclusion

This paper analyzes the origin and persistence of identity-driven choice behavior by modelling decision making under uncertainty in a social context. The model sheds light on the instrumental value of social identity cues for agents that are only able to imperfectly learn about their ability type, and that have to create an estimate of the probability of success of a task for which their individual ability matters. Although these social identity cues are informationally irrelevant in decision making, the model shows that agents can use these cues to optimally manage the over- or under-confidence present in their noisy estimate of ability. The model tells therefore the following story. We are often not sure whether we are 'good enough' to successfully complete a *Competitive* task. Because we have access to a sample of successful people and we can observe several traits of these successful people, we can convince ourselves that there exist correlations between success and certain observable characteristics. The model shows that, if we attribute success or failure to observable characteristics in an optimal way, then this would enable us to form beliefs that would improve our decision making on average. Even though it is unlikely that people are always able to optimally use their social identity cues, the potential value of these cues gives a foundation for why people let their social identity play a role in decision making, even when it is irrelevant. This also implies that the social identities that become salient in a choice context are those that, through the social context, can potentially have instrumental value in the decision-making process.

The option to improve decision making using these social identity cues is not available to all agents. At the aggregate level, this asymmetry can cause choice behavior that induces the existence of a population equilibrium in which the use of social identity cues and the allocation over tasks of individuals belonging to a priori identical subgroups is different. This type of population equilibrium can especially arise when we consider *Competitive* tasks at which only a small fraction of the population succeeds. Moreover, a population equilibrium in which a certain subgroup is overrepresented among the successful individuals at the *Competitive* task exists for a larger set of values when this subgroup represents a majority of the population. Finally, the stronger agents react to their social context, the larger the asymmetry observed in equilibrium will be. Therefore, even when agents manage to behave optimally from an individual perspective, asymmetries in choice behavior across a priori identical subgroups can persist in equilibrium.

These results imply that choice behavior driven by observable characteristics is less likely to be caused by the explicit preferences of agents with these observable characteristics, but is rather driven by the social context in which these agents make their decisions. This supports the results of Nosek et al. (2009), who show that state-level variations in gender stereotypes are a better predictor of the choice of women into STEM careers than SAT scores. This also suggests that any proscribed behavior that follows from social norms (Akerlof and Kranton, 2000) is unlikely to be driven by the internal manifestations of this identity, but is rather an adaptation to what people see around them. The model predicts therefore that choice behavior is, at least partially, socially constructed. This implies that, when we want to eliminate asymmetries across a priori identical subgroups with different social identities, taking care of discrimination, skill-differences or social pressure is not enough. We have to take into account the biases in decision-making that are induced by the social context, biases that may help people at the individual level, but are not beneficial for society at the aggregate level.

An important limitation of this model is that it assumes homogeneity in both the information agents perceive in the social context and in the way in which agents process this information. People's individual-specific social networks could nevertheless play an important role in the formation of perceptions and this could lead to heterogenous perceptions of the social context across agents. Moreover, this heterogeneity in perceptions could be correlated with observable characteristics through variables such as income, neighbourhood or education. An interesting way in which this model could therefore be extended is by introducing heterogeneity in the perceptions of the social identity cues and analyze what the effect of this heterogeneity is on the equilibrium results.

References

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

D. Austen-Smith and R.J. Fryer. An economic analysis of 'acting white'. *Quarterly Journal of Economics*, 120(2):551–583, 2005.

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

H. Battu, M. McDonald, and Y. Zenou. Oppositional identities and the labor market. *Journal of Population Economics*, 20:643–667, 2007.

R. Benabou and J. Tirole. Self-confidence and personal motivation. *The Quarterly Journal of Economics*, 117(3):871–915, 08 2002.

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

J. Benhabib, Alberto Bisin, and M. Jackson. *Handbook of Social Economics*, chapter 5, pages 201–213. North-Holland, 2010. Theories fo Statistical Discrimination and Affirmative Action: A Survey.

A. Bisin, E. Patacchini, T. Verdier, and Y. Zenou. Formation and persistence of oppositional identities. *European Economic Review*, 55(8):1046–1071, 2011.

P. Bordalo, N. Gennaioli, and A. Shleifer. Stereotypes. *Quarterly Journal of Economics*, 131(4):1753–1794, 2016.

L. Bursztyn and R. Jensen. How does peer pressure affect educational investment. *The Quarterly Journal of Economics*, 130:1329–1367, 2015.

T. Buser, M. Niederle, and H. Oosterbeek. Gender, competitiveness and career choices. *The Quarterly Journal of Economics*, 129(3):1409–1447, 2014.

L. Busztyn and R. Jensen. Social image and economic behavior in the field: Identifying, understanding and shaping social pressure. *Annual Review of Economics*, 9:131–153, 2017.

P. Carneiro, J.J. Heckman, and D.V. Masterov. Labor market discrimination and racial differences in premarket factors. Working paper, University College London, 2003.

K. Chung. Role models and arguments for affirmative action. *American Economic Review*, 90(3):640–648, 2000.

S. Coate and G.C. Loury. Will affirmative-action policies eliminate negative stereotypes? *The American Economic Review*, 83(5):1220–1240, 1993.

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

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

J. Denrell. Vicarious learning, undersampling of failure, and the myths of management. *Organizational Science*, 14(3):227–351, 2003.

C.L. Exley and J.B. Kessler. The gender gap in self-promotion. Technical report, NBER, 2019.

J.A. Flory, Leibbrandt A., and J.A. List. Do competitive work places deter female workers? a large-scale natural field experiment on gender differences in job-entry decisions. Working Paper 16546, NBER, 2010.

K. Hoff and J.E. Stiglitz. Equilibrium fictions: A cognitive approach to societal rigidity. *American Economic Review: Papers and Proceedings*, 100:141–146, 2010.

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.

D. Kahneman and D. Lovallo. Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management Science*, 39(1):1–133, 1993.

D. Kahneman, P. Wakker, and R. Sarin. Back to bentham? explorations of experienced utility. *Quarterly Journal of Economics*, 112:375–407, 1997.

H.H. Kelley. *Attribution in Social Interaction.* General Learning Press, 1971.

M. Lalanne and P. Seabright. The old boy network: Gender differences in the impact of social networks on remuneration in top executive jobs. Working paper, Toulouse School of Economics, 2011.

R. LeBoeuf, E. Shafir, and J. Belyavski Bayuk. The conflicting choices of alternating selves. *Organizational Behavior and Human Decision Processes*, 111(1):48–61, 2010.

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

M. Niederle and L. Vesterlund. Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics*, 122(3):1067–1101, 2007.

B. Nosek, F.L. Smyth, and N. Sriram. National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences*, 106(26):10593–10597, 2009.

M.J. Osborne and A. Rubinstein. Games with procedurally rational players. *American Economic Review*, 88:834–847, 1998.

M. Piccione and A. Rubinstein. On the interpretation of decision problems with imperfect recall. *Games and Economic Behavior*, 20:3–24, 1997.

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.

M. Shayo. Social identity and economic policy. *Annual Review of Economics*, 12:355–389, 2020.

D. K. Sherman and G. L. Cohen. The psychology of self-defense: Self-affirmation theory. *Advances in Experimental Social Psychology*, 38:183–242, 2006.

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

Appendix 1: Motivation for the Belief Formation Processes

In this Appendix, I show how the family of belief formation processes,

$$\hat{p}_{i,C} = \begin{cases} \hat{\alpha}_i & \text{if } \sigma_i = R \\ \eta(\pi_{\theta_i} - p_{\theta_i})\hat{\alpha}_i & \text{if } \sigma_i = NR \end{cases} \tag{18}$$

can be motivated with a subjective Bayesian model. Let $Y_{i,C} = 1$ be the event that the agent is successful at the *Competitive* task and let $O_i = 1$ be the event that the agent's subgroup is overrepresented among the successful individuals, meaning $\pi_{\theta_i} > p_{\theta_i}$, while $O_i = 0$ when $\pi_{\theta_i} = p_{\theta_i}$ and $O_i = -1$ when $\pi_{\theta_i} < p_{\theta_i}$. Agents receive two signals, namely $\hat{\alpha}_i$ and O_i. When agents *Repress* the use of the social identity cue, they have the following model in mind,

$$p_{i,C} = \psi^R(\alpha_i) \tag{19}$$

where α_i is the individual-specific ability type and $\psi^R : \alpha_i \to [0, 1]$. Upon receiving their noisy estimate, agents believe that they are observing $\hat{\alpha}_i = \psi(\alpha_i) = p_{i,C}$. This model implies that O_i is not correlated with $p_{i,C}$. Therefore, if agents receive this signal, they will disregard it and $\hat{p}_{i,C} = \hat{\alpha}_i$.

When agents choose to *Not Repress* the use of the social identity cue, they have the following model in mind,

$$p_{i,C} = \psi^{NR}(\alpha_{\theta_i}, \alpha_i) \tag{20}$$

where α_{θ_i} is the average performance at the *Competitive* task of the agent's subgroup, α_i is the individual-specific ability type and $\psi^{NR} : \alpha_{\theta_i}, \alpha_i \to [0, 1]$. The signal O_i is correlated with α_{θ_i} and therefore contains information about $p_{i,C}$. Consequently, we can write

$$\hat{p}_{i,C} \equiv P(Y_{i,C} = 1|O_i) = P(Y_{i,C} = 1)\frac{P(O_i|Y_{i,C} = 1)}{P(O_i)} \tag{21}$$

$$= \hat{\alpha}_i \eta(\pi_{\theta_i} - p_{\theta_i}) \tag{22}$$

which shows that $\hat{\alpha}_i = P(Y_{i,C} = 1)$ can be interpreted as a prior belief on being successful at the *Competitive* task and $\eta(\pi_{\theta_i} - p_{\theta_i}) = \frac{P(O_i|Y_{i,C}=1)}{P(O_i)}$. A value of $\eta > 1$ when $O_i = 1$ implies that $\frac{P(O_i=1|Y_{i,C}=1)}{P(O_i)} > 1$, which coincides with a model in which the agent is more likely to observe that her group is overrepresented when she herself is successful than when she herself is not successful. Consequently, a value of $\eta < 1$ when $O_i = -1$ implies $\frac{P(O_i=-1|Y_{i,C}=1)}{P(O_i)} < 1$ and coincides with a similar model in which the agent is less likely to observe that her group is underrepresented when she herself is successful than when she herself is not successful. Finally, a value $\eta = 1$ when $O_i = 0$ coincides with a model in which it is equally likely to observe $\pi_{\theta_i} = p_{\theta_i}$ when the agent herself is successful as when she herself is not successful. In this case, $\hat{p}_{i,C} = \hat{\alpha}_i$ and the strategies *Repress* and *Not Repress* are equivalent.

Appendix 2: Proofs

PROPOSITION 1 (The Optimal Use of Social Identity): *The optimal strategies σ^* given an agent's type $\{\alpha, \theta\}$ are the following:*

- *For agent's of type $\{\alpha, \theta\}$ such that $\alpha > \gamma$ and $\pi_\theta > \frac{1}{2}$ or $\alpha < \gamma$ and $\pi_\theta < \frac{1}{2}$, $\sigma^* = NR$*
- *For agent's of type $\{\alpha, \theta\}$ such that $\alpha > \gamma$ and $\pi_\theta < \frac{1}{2}$ or $\alpha < \gamma$ and $\pi_\theta > \frac{1}{2}$, $\sigma^* = R$*

Proof. Agents choose σ_i to maximize V_i. Consider first agents that have an ability type such that $\alpha > \gamma$. The welfare-maximizing choice for these agents is to take action $a = C$. Therefore, V_i is larger when playing NR than when playing R if and only if the probability that these agents choose $a = C$ is larger when choosing $\sigma_i = NR$ than when choosing $\sigma_i = R$, which means $\Phi_{\alpha,\theta,NR,\Pi} \geq \Phi_{\alpha,\theta,R,\Pi}$. Since $\Phi_{\alpha,\theta,\sigma,\Pi} = P(\hat{\alpha} > \gamma_\sigma|\alpha)$, this is the case when $\gamma_{NR} < \gamma_R$. This is true if and only if $\pi_\theta \geq \frac{1}{2}$. Therefore, NR is only an optimal strategy for agents with $\alpha > \gamma$ when their observable characteristic θ is such that the social identity cue $\pi_\theta \geq \frac{1}{2}$. If this is not the case, they are better off choosing strategy $\sigma_i = R$. Vice versa for agents with $\alpha < \gamma$, the welfare-maximizing choice is to take action $a = NC$. Therefore, V_i i larger when playing NR than when playing R if and only if the probability that these agents choose $a = NC$ is larger when choosing $\sigma_i = NR$ than when choosing $\sigma_i = R$, which means $\Phi_{\alpha,\theta,NR,\Pi} \leq \Phi_{\alpha,\theta,R,\Pi}$. This is the case if and only if $\gamma_{NR} > \gamma_R$, meaning that we need $\pi_\theta \leq \frac{1}{2}$. Therefore, agents with α should only choose strategy $\sigma_i = NR$, when their observable characteristic is such that $\pi_\theta \leq \frac{1}{2}$. Otherwise, they are better off choosing strategy $\sigma_i = R$. ■

PROPOSITION 2: *Assume $p_0 = p_1$. A population equilibrium with a 'Neutral Regime' always exists and can co-exist with a stable population equilibrium with a 'Non-Neutral Regime' when the following condition is satisfied,*

$$\gamma > \frac{2}{\epsilon_{S,\gamma}(0)\frac{\partial \eta(0)}{\partial \pi_\theta}} \tag{23}$$

where S is the total number of individuals that have successfully completed the Competitive task and $\epsilon_{S,\gamma}(0) = -\frac{\frac{\partial S}{\partial \gamma}|_{\pi_\theta = p_\theta}}{S}$.

Proof. Let $p_1 = p_0$ and assume WLOG that $\pi_1 > \pi_0$. Proposition 1 shows that all agents with $\theta = 1$ and $\alpha > \gamma$ will choose $\sigma_i = NR$, while all agents with $\theta = 0$ and $\alpha > \gamma$ will choose $\sigma_i = R$. Similarly, all agents with $\theta = 1$ and $\alpha < \gamma$ will choose $\sigma_i = R$, while all agents with $\theta = 0$ and $\alpha < \gamma$ will choose $\sigma_i = NR$. The social identity cue $\tilde{\pi}_1(\pi_1, \sigma)$ that is induced by the collection of strategies σ and π_1 is given by,

$$\tilde{\pi}_1(\pi_1, \sigma) = \frac{S_1}{S_1 + S_0} \tag{24}$$

where $S_1 = \int \alpha \Phi_{\alpha,1,\sigma,\Pi} f(\alpha) d\alpha$ and $S_0 = \int \alpha \Phi_{\alpha,0,\sigma,\Pi} f(\alpha) d\alpha$. First of all, one can directly infer that $\pi_1 = \frac{1}{2}$ is an equilibrium, because in that case, the strategies NR and R are equivalent. Since α and θ are independently distributed, this implies that $S_1 = S_0$ and therefore $\tilde{\pi}_1(\frac{1}{2}, \sigma) = \frac{1}{2}$.

Furthermore, one can show that S_1 is bounded from above. Let $\overline{S}_1$ be the upper bound on S_1. This upper bound arises in the extreme case in which all agents with $\theta = 1$ and $\alpha > \gamma$ choose to undertake the *Competitive* task, and is equal to

$$\overline{S}_1 = \int_{\alpha<\gamma} \int_{\hat{\alpha}>\gamma} \alpha g_\alpha(\hat{\alpha}) f(\alpha) d\hat{\alpha} d\alpha + \int_{\alpha>\gamma} \alpha f(\alpha) d\alpha \tag{25}$$

Similarly, one can show that S_0 is bounded from below. Let $\underline{S}_0$ be the lower bound on S_0. This lower bound arises in the extreme case in which no agents with $\theta = 0$ and $\alpha < \gamma$ choose to undertake the *Competitive* task, and is equal to

$$\underline{S}_0 = \int_{\alpha>\gamma} \int_{\hat{\alpha}>\gamma} \alpha g_\alpha(\hat{\alpha}) f(\alpha) d\hat{\alpha} d\alpha \tag{26}$$

Consequently, $\tilde{\pi}_1(\pi_1, \sigma)$ has an upper bound, that is given by

$$\overline{\pi}_1(\pi_1, \sigma) = \frac{\overline{S}_1}{\overline{S}_1 + \underline{S}_0} < 1 \tag{27}$$

It is now sufficient to show that, if $\frac{\partial \tilde{\pi}_1(\pi_1,\sigma)}{\partial \pi_1}|_{\pi_1=\frac{1}{2}} > 1$, then there exists a stable 'Non-Neutral Regime'.

Let $S = S_0 + S_1$, such that $\tilde{\pi}_1(\pi_1, \sigma) = \frac{S_1}{S}$. Then,

$$\frac{\partial \tilde{\pi}_1(\pi_1, \sigma)}{\partial \pi_1}|_{\pi_1=\frac{1}{2}} = \frac{\frac{\partial S_1}{\partial \pi_1} S - \frac{\partial S}{\partial \pi_1} S_1}{S^2} \tag{28}$$

$$= \frac{\frac{\partial S_1}{\partial \pi_1} - \frac{\partial S}{\partial \pi_1}\frac{1}{2}}{S} \tag{29}$$

$$= \frac{\frac{1}{2}\left[\frac{\partial S_1}{\partial \pi_1} - \frac{\partial S_0}{\partial \pi_1}\right]}{S} \tag{30}$$

Therefore, $\frac{\partial \tilde{\pi}_1(\pi_1,\sigma)}{\partial \pi_1}|_{\pi_1=\frac{1}{2}} > 0$ if and only if $\frac{\partial S_1}{\partial \pi_1} > \frac{\partial S_0}{\partial \pi_1}$. $\frac{\partial S_1}{\partial \pi_1}$ for $\pi_1 \in [\frac{1}{2}, 1]$ is given by,

$$\begin{aligned}
\frac{\partial S_1}{\partial \pi_1} &= \frac{\partial}{\partial \pi_1} \int \alpha \Phi_{\alpha,1,\sigma,\Pi} f(\alpha) d\alpha \\
&= \frac{\partial}{\partial \pi_1} \left(\int_{\alpha>\gamma} \alpha \Phi_{\alpha,1,NR,\Pi} f(\alpha) d\alpha + \int_{\alpha<\gamma} \alpha \Phi_{\alpha,1,R,\Pi} f(\alpha) d\alpha \right) \\
&= \frac{\partial}{\partial \pi_1} \left(\int_{\alpha>\gamma} \int_{\hat{\alpha}>\frac{\gamma}{\eta_1}} \alpha g_\alpha(\hat{\alpha}) f(\alpha) d\hat{\alpha} d\alpha + \int_{\alpha<\gamma} \int_{\hat{\alpha}>\gamma} \alpha g_\alpha(\hat{\alpha}) f(\alpha) d\hat{\alpha} d\alpha \right) \\
&= \int_{\alpha>\gamma} \alpha g_\alpha \left(\frac{\gamma}{\eta_1}\right) \frac{\gamma}{\eta_1^2} \frac{\partial \eta(\pi_1 - p_1)}{\partial \pi_1} f(\alpha) d\alpha
\end{aligned}$$

where $\eta_1 = \eta(\pi_1 - p_1)$. Similarly,

$$\frac{\partial S_0}{\partial \pi_1} = -\int_{\alpha<\gamma} \alpha g_\alpha \left(\frac{\gamma}{\eta_0}\right) \frac{\gamma}{\eta_0^2} \frac{\partial \eta(\pi_1 - p_1)}{\partial \pi_1} f(\alpha) d\alpha$$

where $\eta_0 = \eta(\pi_0 - p_0)$. The fact that $\frac{\partial S_1}{\partial \pi_1}$ is positive is not surprising, and reflects the fact that, when π_1 increases, more agents with $\alpha > \gamma$ and $\theta = 1$ will choose the competitive task. The fact that $\frac{\partial S_0}{\partial \pi_1}$ is negative is neither surprising, and reflects the fact that, when π_1 increases, π_0 decreases. Consequently, more agents with $\alpha < \gamma$ and $\theta = 0$ will choose the non-competitive task and S_0 decreases.

When we evaluate both partial derivatives at $\pi_1 = \frac{1}{2}$, we have $\eta_0 = \eta_1 = 1$ and we can write,

$$\frac{\partial \tilde{\pi}_1(\pi_1, \sigma)}{\partial \pi_1}|_{\pi_1 = \frac{1}{2}} = \frac{\frac{1}{2}\frac{\partial \eta(0)}{\partial \pi_1}\gamma \int \alpha g_\alpha(\gamma) f(\alpha) d\alpha}{S} \tag{31}$$

Furthermore, S evaluated at $\pi_1 = \frac{1}{2}$ gives us,

$$S|_{\pi_1 = \frac{1}{2}} = \int \int_{\hat{\alpha} > \gamma} \alpha g_\alpha(\hat{\alpha}) f(\alpha) d\hat{\alpha} d\alpha \tag{32}$$

and consequently,

$$\frac{\partial S}{\partial \gamma}|_{\pi_1 = \frac{1}{2}} = -\int \alpha g_\alpha(\gamma) f(\alpha) d\alpha \tag{33}$$

which is always negative and reflects the fact that, the higher the threshold γ, the lower the total number of agents that will choose to undertake the *Competitive* task. Let $\epsilon_{S,\gamma}(0) = -\frac{\frac{\partial S}{\partial \gamma}|_{\pi_1 = p_1}}{S}$. We can then write,

$$\frac{\partial \tilde{\pi}_1(\pi_1, \sigma)}{\partial \pi_1}|_{\pi_1 = \frac{1}{2}} = \epsilon_{S,\gamma}(0) \frac{1}{2} \frac{\partial \eta(0)}{\partial \pi_1} \gamma \tag{34}$$

Since a stable non-neutral equilibrium exists when $\frac{\partial \tilde{\pi}_1(\pi_1, \sigma)}{\partial \pi_1}|_{\pi_1 = \frac{1}{2}} > 1$, we arrive to the following condition

$$\gamma > \frac{2}{\epsilon_{S,\gamma}(0) \frac{\partial \eta(0)}{\partial \pi_1}} \tag{35}$$

■

COROLLARY 1: *Let $\theta' \in \Theta$ be the complement of θ and assume WLOG that $\pi_\theta > \pi_{\theta'}$. The optimal use of social identity has both a population effect, such that $\Phi_{\alpha,\theta,\sigma_i,\Pi} > \Phi_{\alpha,\theta',\sigma_i,\Pi}$ and a selection effect, such that $E(\alpha|\hat{p}^{\sigma_i} > \gamma, \theta) < E(\alpha|\hat{p}^{\sigma_i} > \gamma, \theta')$. The strength of both effects is such that the order $\pi_\theta > \pi_{\theta'}$ will always be preserved.*

Proof. Assume WLOG that $\pi_1 > \pi_0$. Then, $\frac{\gamma}{\eta(\pi_1 - p_1)} < \gamma$, while $\frac{\gamma}{\eta(\pi_0 - p_0)} > \gamma$. Therefore, all agents with $\alpha > \gamma$ and $\theta = 1$ will choose $\gamma_i = \frac{\gamma}{\eta(\pi_1 - p_0)}$, while all agents with $\alpha > \gamma$ and $\theta_i = 0$ will choose $\gamma_i = \gamma$. Consequently, $\Phi_{\alpha,1,NR,\Pi} > \Phi_{\alpha,0,R,\Pi}$ for all α. Because N is arbitrarily large, these probabilities can be interpreted as population fractions, which proves the population effect. Furthermore, because $\frac{\gamma}{\eta(\pi_1 - p_0)} < \gamma$, we know that agents with $\theta = 0$ will choose the *Competitive* task for on average higher realizations of $\hat{\alpha}$. Because $E(\hat{\alpha}) = \alpha$, these agents will on average also have higher true ability levels, which leads to the selection effect $E(\alpha|\hat{p}^{NR} > \gamma, 1) < E(\alpha|\hat{p}^{R} > \gamma, 0)$. Finally, in Proposition 2 we have shown that,

$$\frac{\partial \tilde{\pi}_1(\pi_1, \sigma)}{\partial \pi_1} = \frac{\pi_1 \left[\frac{\partial S_1}{\partial \pi_1} - \frac{\partial S_0}{\partial \pi_1}\right]}{S} \tag{36}$$

and $\frac{\partial S_1}{\partial \pi_1} > 0$, while $\frac{\partial S_0}{\partial \pi_1} < 0$. Therefore, $\frac{\partial \tilde{\pi}_1(\pi_1,\sigma)}{\partial \pi_1} > 0$ and the selection and population effect will not reverse the order $\pi_1 > \pi_0$. ■

COROLLARY 2: *The set of values γ and reaction functions η for which the sufficient condition of Proposition 2 for the existence of a population equilibrium with a 'Non-Neutral Regime' in which $\pi_1 > p_0$ is satisfied increases when $p_1 > p_0$.*

Proof. We can adjust the proof of Proposition 2 by correcting for the fact that $p_1 \neq p_0$. The social identity cue $\tilde{\pi}_1(\pi_1, \sigma)$ that is induced by the collection of strategies σ and π_1 is given by,

$$\tilde{\pi}_1(\pi_1, \sigma) = \frac{S_1}{S_1 + S_0} \tag{37}$$

where $S_1 = p_1 \int \alpha \Phi_{\alpha,1,\sigma,\Pi} f(\alpha) d\alpha$ and $S_0 = p_0 \int \alpha \Phi_{\alpha,\theta,\sigma,\Pi} f(\alpha) d\alpha$. One can again directly infer that $\pi_1 = p_1$ is an equilibrium and the proof of the existence of a 'Non-Neutral' regime goes therefore along the same lines. Specifically,

$$\frac{\partial \tilde{\pi}_1(\pi_1, \sigma)}{\partial \pi_1}|_{\pi_1 = p_1} = \frac{p_1 \left[\frac{\partial S_1}{\partial \pi_1} - \frac{\partial S_0}{\partial \pi_1}\right]}{S} \tag{38}$$

where,

$$\frac{\partial S_1}{\partial \pi_1} = p_1 \int_{\alpha > \gamma} \alpha g_\alpha \left(\frac{\gamma}{\eta_1}\right) \frac{\gamma}{\eta_1^2} \frac{\partial \eta(\pi_1 - p_1)}{\partial \pi_1} f(\alpha) d\alpha$$

and,

$$\frac{\partial S_0}{\partial \pi_1} = -p_0 \int_{\alpha<\gamma} \alpha g_\alpha \left(\frac{\gamma}{\eta_0}\right) \frac{\gamma}{\eta_0^2} \frac{\partial \eta(\pi_1 - p_1)}{\partial \pi_1} f(\alpha) d\alpha$$

When we evaluate both these partial derivatives at $\pi_1 = p_1$, we get,

$$\frac{\partial \tilde{\pi}_1(\pi_1, \sigma)}{\partial \pi_1}|_{\pi_1=p_1} = \frac{p_1 \frac{\partial \eta(0)}{\partial \pi_1} \gamma [p_1 \int_{\alpha>\gamma} \alpha g_\alpha(\gamma) f(\alpha) d\alpha + p_0 \int_{\alpha<\gamma} \alpha g_\alpha(\gamma) f(\alpha) d\alpha]}{S} \tag{39}$$

Along the lines of the proof of Proposition 2, we can show that a 'Non-Neutral' regime exists when $\frac{\partial \tilde{\pi}_1(\pi_1,\sigma)}{\partial \pi_1}|_{\pi_1=p_1} > 1$. This gives the following condition,

$$\gamma > \frac{S}{\frac{\partial \eta(0)}{\partial \pi_1} p_1 [p_1 \Delta_{\alpha>\gamma} + p_0 \Delta_{\alpha<\gamma}]} \tag{40}$$

where $\Delta_{\alpha<\gamma} = \int_{\alpha<\gamma} \alpha g_\alpha(\gamma) f(\alpha) d\alpha$ and $\Delta_{\alpha>\gamma} = \int_{\alpha>\gamma} \alpha g_\alpha(\gamma) f(\alpha) d\alpha$. Because $p_1[p_1 \Delta_{\alpha>\gamma} + p_0 \Delta_{\alpha<\gamma}]$ is increasing in p_1, the threshold for γ decreases when p_1 increases. ■

COROLLARY 3: *Take two reaction functions $\hat{\eta}$ and η, such that $\hat{\eta}(\pi_\theta - p_\theta) > \eta(\pi_\theta - p_\theta)$ for all $\pi_\theta > p_\theta$. Assume WLOG that a 'Non-Neutral Regime' exists in which $\pi_1 > p_1$. Let $\pi^*_{\eta,1}$ be the equilibrium value of π_1 given a reaction function η. Then, $\pi^*_{\hat{\eta},1} > \pi^*_{\eta,1}$.*

Proof. Assume $\eta(\pi_\theta - p_\theta)$ is a reaction function such that, given a value of γ, the sufficient condition for the existence of a 'Non-Neutral Regime' of the form $\pi_1 > p_1$ and $\pi_0 < p_0$ is satisfied. To proof Corollary 3, we need to proof first of all that for any reaction function $\hat{\eta}(\pi_\theta - p_\theta)$, such that $\hat{\eta}(\pi_1 - p_1) > \eta(\pi_1 - p_1)$ for all $\pi_1 > p_1$, a 'Non-Neutral Regime' exists. If we assume that a 'Non-Neutral Regime' exists for a reaction function $\eta(\pi_\theta - p_\theta)$, then it follows from Proposition 2 that a 'Non-Neutral Regime' also exists for any reaction function $\hat{\eta}(\pi_\theta - p_\theta)$. Furthermore, let $\tilde{\pi}_{\eta,1}(\pi_1, \sigma)$ be the induced value of π_1 for a reaction function η. Then, if $\hat{\eta}(\pi_1 - p_1) > \eta(\pi_1 - p_1)$ for all $\pi_1 > p_1$,

$$\tilde{\pi}_{\hat{\eta},1}(\pi_1, \sigma) > \tilde{\pi}_{\eta,1}(\pi_1, \sigma) \quad \forall \pi_1 > p_1 \tag{41}$$

Consequently, let $\pi^*_{\eta,1}$ be the equilibrium value of π_1 that arises in a 'Non-Neutral

Regime' for a reaction function η. Then,

$$\pi^{(1)} \equiv \tilde{\pi}_{\hat{\eta},1}(\pi^*_{\eta,1}, \sigma) > \tilde{\pi}_{\eta,1}(\pi^*_{\eta,1}, \sigma) = \pi^*_{\eta,1} \tag{42}$$

which implies that,

$$\pi^{(2)} \equiv \tilde{\pi}_{\hat{\eta},1}(\pi^{(1)}, \sigma) > \tilde{\pi}_{\hat{\eta},1}(\pi^*_{\eta,1}, \sigma) \equiv \pi^{(1)} \tag{43}$$

and,

$$\pi^{(3)} \equiv \tilde{\pi}_{\hat{\eta},1}(\pi^{(2)}, \sigma) > \tilde{\pi}_{\hat{\eta},1}(\pi^{(1)}, \sigma) \equiv \pi^{(2)} \tag{44}$$

This sequence converges to $\pi^*_{\hat{\eta},1} = \tilde{\pi}_{\hat{\eta},1}(\pi^*_{\hat{\eta},1}, \sigma)$ and is everywhere above $\pi^*_{\eta,1}$ and below the upper bound $\overline{\pi}_1$ on π_1. This shows that, for any reaction function $\hat{\eta}(\pi_\theta - p_\theta)$ such that $\hat{\eta}(\pi_1 - p_1) > \eta(\pi_1 - p_1)$ for all $\pi_1 > p_1$, in equilibrium

$$\pi^*_{\hat{\eta},1} > \pi^*_{\eta,1} \tag{45}$$

■

COROLLARY 4: *Assume WLOG that agents hold an incorrect belief $\hat{p}_\theta > p_\theta$. In that case, a population equillibrium with a 'Neutral Regime' no longer exists and there will only exist a population equilibrium with a 'Non-Neutral Regime' in which $\pi_\theta < p_\theta$.*

Proof. Assume WLOG that $\hat{p}_0 > p_0$. This means that,

$$\eta(\pi_0 - p_0) = \begin{cases} > 1 & \text{if } \pi_0 > \hat{p}_0 \\ 1 & \text{if } \pi_0 = \hat{p}_0 \\ < 1 & \text{if } \pi_0 < \hat{p}_0 \end{cases} \tag{46}$$

and consequently, when $\pi_0 = p_0$, $\eta(\pi_0 - p_0) < 1$. This implies that $\tilde{\pi}_0(p_0, \sigma) < p_0$ and $\pi_0 = p_0$ is not an equilibrium. Furthermore, because $\eta(\pi_0 - p_0) < 1$ implies $\eta(\pi_1 - p_1) > 1$, it follows that $\tilde{\pi}_1(p_1, \sigma) > p_1$. As shown in the proof of Proposition 2, $\tilde{\pi}_1$ is bounded from above and therefore there exists a population equilibrium with a 'Non-Neutral Regime' in which $\pi_0 < p_0$. ■