CAMBRIDGE WORKING PAPERS IN ECONOMICS
JANEWAY INSTITUTE WORKING PAPERS

On the Origin and Persistence of Identity-Driven Choice Behavior

Caroline W.
Liqui Lung
University of Cambridge

Abstract

Recent literature shows how a priori identical individuals belonging to different social groups make different choices. This paper proposes a novel explanation for this identity-driven choice behavior. Agents choose whether to undertake a task with a probability of success driven by an ability. They have a noisy perception of this ability and observe social cues that stem from the prevalence of their subgroup among the successful individuals. Although the noise in their perception is unbiased, it has an asymmetric effect on expected utility. This makes it optimal for certain agents to bias their noisy perception with social cues, even when these cues are irrelevant in a Bayesian sense. I show the existence of a stable population equilibrium in which both task allocation and the use of social cues differ between a priori identical subgroups.

Reference Details

2274	Cambridge Working Papers in Economics
2232	Janeway Institute Working Paper Series
Published	15 December 2022
Revised	4 January 2023
JEL codes	D81, D91, I240, Z13
Websites	www.econ.cam.ac.uk/cwpe
	www.janeway.econ.cam.ac.uk/working-papers

On the Origin and Persistence of Identity-Driven Choice Behavior

Caroline W. Liqui Lung *

Version: January 4, 2023

Recent literature shows how a priori identical individuals belonging to different social groups make different choices. This paper proposes a novel explanation for this identity-driven choice behavior. Agents choose whether to undertake a task with a probability of success driven by an ability. They have a noisy perception of this ability and observe social cues that stem from the prevalence of their subgroup among the successful individuals. Although the noise in their perception is unbiased, it has an asymmetric effect on expected utility. This makes it optimal for certain agents to bias their noisy perception with social cues, even when these cues are irrelevant in a Bayesian sense. I show the existence of a stable population equilibrium in which both task allocation and the use of social cues differ between a priori identical subgroups.
(JEL: D-81, D-91, I-240, Z-13)

*University of Cambridge (Email: cll64@cam.ac.uk). I would like to thank Olivier Compte, Philippe Jehiel, Roland Bénabou, Debraj Ray, Alessandra Casella, Abhijit Banerjee, Muhamet Yildiz, Vincent Crawford, Willemien Kets, Dilip Mookherjee, Gabrielle Fack, Jean-Marc Tallon and the participants of SITE, ERINN, ThReD, the Royal Economic Society Annual Conference, the North American, Asian and China Meetings of the Econometric Society, ADRES, the ECINEQ annual conference, the Warwick Economics PhD Conference and the seminars at MIT, PSE, the University of Cambridge and the University of Amsterdam (CREED) for their useful comments and suggestions.

1 Introduction

Equally performing individuals belonging to different social groups can have systematically different beliefs about their own abilities. Recent studies have found evidence for this phenomenon for male and female math students (Lippmann and Senik, 2018), black and white college students (Hodge et al., 2008), men and women in the labor market (Exley and Kessler, 2022) and socially more and less advantaged children in high school (Guyon and Huillery, 2021), amongst others. Such differences in beliefs can be driven by the social context in which these beliefs are formed, even when the information that can be retrieved from this context is irrelevant in the belief formation process.[1] Furthermore, these differences in beliefs are one of the main drivers of identity-driven choice behavior, meaning a priori identical individuals belonging to different social groups making different choices. This contributes to inequalities across social groups and negatively affects social diversity in occupational and educational fields. It is therefore important to understand how such differences in beliefs arise. This paper contributes to the literature by showing why agents may find it optimal to let their beliefs be influenced by irrelevant statistics about who is successful, and how this can lead to persistent differences in choice behavior across a priori identical subgroups. I discuss how the insights this paper provides can be useful to develop the appropriate informational policies needed to achieve social diversity and fight harmful stereotypes.

I propose a model in which the effect of social context on belief formation is determined endogenously, without assuming social context directly affects utility. To analyze the origin of identity-driven choice behavior, I analyze agents that attempt to form a subjective belief about the individual-specific probability of success in a task. In this process, they choose whether they let irrelevant statistics about the prevalence of their social group among the successful individuals in the task influence their belief formation. These subjective beliefs induce choice behavior, and this choice behavior leads to successes and failures that in turn give rise to statistics. To analyze the persistence of identity-driven choice behavior, I use a novel static solution concept that tractably captures the mutually stable choices of belief formation and tasks in this process.

In the model, agents choose between an *ability-driven* task with an individual-specific probability of success and an outside option with a probability of success that is the same for all agents. Examples of *ability-driven* tasks can be math-related tasks or leadership-related tasks. The key assumption in the model is that agents only have a noisy perception of their probability of success in this *ability-driven* task[2]. This assumption aims to

[1] See for example Lippmann and Senik (2018), Coffman (2014) and Flory et al. (2010)

[2] I assume agents are not systematically over- or under confident. This assumption only serves to show

capture the fact that, driven by exogenous factors such as emotions, agents can sometimes be a bit too optimistic about their chances of success, while other times they may be too pessimistic (see e.g. Fiedler and Bless (2000) and Elster (1996)). Agents attempt to form a subjective belief about their probability of success in the *ability-driven* task and derive utility from successfully completing either option. As is standard in subjective Bayesian models, they choose a task to maximize subjective expected utility.

Agents are also described by an observable characteristic. This characteristic can represent for example an agent's gender, race or social class, and is independently distributed from the individual-specific probability of success in the *ability-driven* task. I envision agents that have an imperfect idea about their environment and think that using social context could be useful to form a belief about their individual-specific probability of success. Consequently, I assume agents observe social identity cues that stem from the prevalence of their subgroup among the successful individuals in the task. In the spirit of the ideas presented in Compte and Postlewaite (2019), agents choose between two subjective belief-formation processes; one in which they naively follow their own noisy perception, and one in which this noisy perception is influenced by their social identity cue. This cue affects belief formation in a direction contingent on the agent's observable type. Specifically, when an agent's social group is relatively overrepresented among the successful individuals in the task, the cue can be used to bias the noisy perception upwards, while vice versa when their social group is relatively underrepresented.

The choices of subjective belief formation induce choices of tasks, and, at the aggregate level, these choices of tasks give rise to social identity cues. To study the mutually stable choices of belief formation and tasks in this dynamic process, I use a static solution concept[3]. Specifically, I define an individual optimality criterion that states a subjective belief formation process is optimal when it maximizes expected utility on average over all possible realizations of the agent's noisy perception. I then analyze the existence and stability of the fixed points in the social identity cues induced by these individually optimal strategies. As in Compte and Postlewaite (2004) and Brunnermeier and Parker (2005), individual optimality implies agents deviate from Bayesian updating when this increases their expected utility. This can be justified with the view that agents learn their optimal belief formation through their own experience with similar choices[4].

that such a systematic bias is not what drives the results in this model. Furthermore, similar results can be derived when the perception of the probability of success in both the *ability-driven* task and in the outside option are noisy.

[3]A similar approach is used in Dekel et al. (2007)

[4]The dynamic story underlying the reduced-form analysis is the following. Agents make similar choices throughout their lifetime of whether to undertake a task with a probability of success that depends on a particular ability. For example, early in life people choose whether to 'enter a math competition' or

The idea that equilibrium is the result of people attempting to choose among strategies according to their fitness value is standard. The non-standard aspects of this model are that the set of strategies represents a limited set of belief-formation processes, and that equilibrium beliefs are disciplined in a manner that is generally absent in equilibrium concepts such as the Berk-Nash Equilibrium (Esponda and Pouzo, 2016), the Self-Confirming Equilibrium (Fudenberg and Levine, 1993) or the Personal Equilibrium (Spiegler, 2016). Specifically, in the latter concepts, equilibrium beliefs are consistent with observational feedback, which ensures subjective beliefs being closest to the truth. Because in this paper beliefs are consistent with fitness, this allows agents to make decisions that are better aligned with welfare maximization. These beliefs are not consistent with Bayesian updating, and the model raises the question whether there is legitimacy for using Bayesian updating as a belief formation rule when perceptions are noisy to start with and agents do not have the tools or instruments to correct for this noisiness.

The behavior implied by the individual optimality criterion allows us to shed light on the origin of identity-driven choice behavior. Specifically, I show that, although the 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. This is driven by the asymmetric effect of the noisiness in the agent's perception on expected utility. When agents with relatively high chances of success in the *ability-driven* task are too pessimistic, this may induce them to mistakenly choose to undertake the outside option. For these agents, biasing their noisy, and possibly over-pessimistic belief upward is therefore welfare improving. For agents with relatively low chances of success, this is the case when biasing their noisy, and possibly too optimistic belief downward. As a result, agents with relatively high chances of success will wish to use their social identity cue when they belong to the socially more successful group, while they will wish to avoid it when they belong to the socially less successful group. The opposite holds for agents with relatively low chances of success. Consequently, when beliefs are noisy, it becomes useful to incorporate cues in the decision function that would otherwise be ignored, and social identity can drive choice behavior, even when it is uninformative about performance and does not directly affect utility[5]

'undertake a math-related major', while later in life they choose whether to 'pursue a career in a STEM field'. Agents experiment with the different belief formation processes and learn what process leads to choices that result in more successful outcomes on average, without knowing the exact relation between their choice of belief formation process, their choice of task and the observed outcomes. Furthermore, because I assume in the solution concept that agents play their individually optimal strategies, this analysis is more suitable for settings that appear later in life, such as occupational choices.

[5]Hence, the optimal strategy is driven by an agent type, of which the agent remains unaware. Besides the earlier described coarse learning from experience, an alternative foundation for this behavior could

The current economic literature models the effect of social identity on decision making through direct utility derived from self-image (Akerlof and Kranton (2000), Benabou and Tirole (2011)) or from the psychological costs of interaction with others stemming from the fear of being punished by peers for not complying with social norms (Akerlof and Kranton, 2000), the fear of being rejected by peers (Austen-Smith and Fryer, 2005), costly interaction with people different from yourself (Battu et al., 2007) or other forms of social pressure[6] This model provides a different, but possibly complementary view, where the use of social identity in belief formation can be a mechanism that gives agents the option to manage confidence. Specifically, the use of social identity cues in belief formation allows agents to manage the degree of over- or under-confidence regarding their chances of success through the distinct processing of their noisy perception. The optimal use of social context in belief formation results in the optimal management of confidence to limit the adverse effects of holding inaccurate beliefs, and therefore, 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), Brunnermeier and Parker (2005) or Benabou and Tirole (2002).

The options to manage confidence available to agents depend on their observable characteristic, and, depending on social context, they can be asymmetric across different social groups[7]. This asymmetry can potentially lead to different choices of tasks across a priori identical subgroups. This could feed differences in the prevalence of a particular subgroup among the successful people, which in turn fuels the differential use of social identity cues in belief formation. I show the existence of a stable population equilibrium in which the task allocation and the use of social identity cues in belief formation differ between a priori identical subgroups. Therefore, even when agents manage to use their social context in an optimal way from an individual perspective, identity-driven choice behavior can persist. Depending on how agents process information, minority groups can react more strongly to changes in social context than majority groups. This can

be in the spirit of Benabou and Tirole (2006), where there are two selves. A calm, rational self knows the true ability type, while a less calm self can be biased in the moment. The first self may then decide on a belief formation rule, while the second self chooses an action at a given point in time, given its noisy perception of ability and the earlier chosen belief formation rule. The rational self ties the hands of the 'in the moment' self to limit its adverse effects on choice. This foundation is also in line with the more popular concept of 'life hacks' (Peters, 2021).

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

[7]This is in 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)). Similarly, Exley and Kessler (2022) 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.

lead to a larger relative under- or overrepresentation of the minority group among the successful individuals compared to the majority group in equilibrium.

Interestingly, the differential use of social identity cues in belief formation induces both a difference in the propensity to choose the *ability-driven* task across subgroups, and a difference in mean competence: for agents belonging to the socially less successful subgroup, choosing the *ability-driven* task requires a higher noisy perception, 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 *ability-driven* 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.[8] This result is in line with Niederle and Vesterlund (2007), who show that too few high-skilled women and too many low-skilled men enter competitive math-related tasks. Finally, I show that the influence of the social context on beliefs especially drives choice behavior of agents with average ability levels, because agents with extreme ability levels are always more likely to make their welfare-maximizing choice, independent of 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 rest of the paper is organized as follows. Section 2 reviews the relevant literature. Section 3 introduces the model. Section 4 presents the results. Section 5 discusses the assumptions of the model. Section 6 presents possible applications, and Section 7 concludes. The formal proofs of the results can be found in Appendix 1.

2 Social Identity and Belief Formation

The idea that social context affects a person's belief formation regarding her own chances of success or abilities through social identity finds its origin in the field of social psychology. Hogg and Grieve (1999) discuss how in the process of depersonalization, which is associated with social identification, individual and concomitant unshared beliefs, attitudes, feelings, and behaviors are replaced by an in-group prototype that prescribes shared beliefs, attitudes, feelings and behaviors. Similarly, Seligman (2006) argues how people can interpret numerous failures from others like them as evidence that they will

[8] 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.

fail as well. Finally, Steele (2010) discusses how the psyche of the individual could get damaged by bad images of the group projected in society. Repeated exposure to these images causes these images to be internalized, which damages character by causing low self esteem, low expectations, low motivation and self doubt.

There is also a large body of evidence showing that social identity affects choice behavior and preferences. For example, Smith et al. (2007) shows that, when people complete a high stereotype-threat test, they report decreased task interest. Davies et al. (2002) argues how the combination of decreased enjoyment and diminished self-confidence explains why women experiencing stereotype threat report less interest in math and science fields and weaker leadership aspirations compared to men or non-threatened women. Similarly, Banaji and Greenwald (2016) show how implicit associations picked up from social context by our automatic brain affect our behavior, such as the intellectual pursuits we select, and Perry et al. (2003) discusses how people tend to protect themselves against stereotype threat by ceasing to care about the domain in which the stereotype applies. LeBoeuf et al. (2010) shows how the choices of the presents we would like to receive depend on the identity made salient by the experimenter. Finally, Oh (2019) shows how Indian workers are willing to forego substantial payments to avoid tasks that are associated with other castes.

Hogg and Grieve (1999) define two classes of motivation for social identification. The first motivation is self-esteem. People are motivated to maintain or achieve positive distinctiveness for their own group relative to other groups, because intergroup evaluation is self-evaluation. This idea is introduced in economics by Akerlof (2016a). He argues that group pride facilitates collective action, and proposes a model in which agents will be inclined to think in we-terms when the group is a source of pride and when agents are individually accorded low self-esteem. Similarly, Akerlof (2016b) proposes a model for the choice of values. In this paper, he argues that these choices result from the conflicting desires of people to be esteemed by their peers, but at the same time wanting to have self-esteem, which is often best satisfied by differentiating from peers.

Hogg and Grieve (1999) argue nevertheless that self-esteem may not be a necessary nor sufficient motivation, and that subjective uncertainty reduction may be the critical driver for social identification. Subjective certainty gives people confidence about how to behave, and what to expect from their physical and social environment. Ben-Ner et al. (2009) shows in an experiment how the importance of a source of identity varies with the type of behavior under consideration, and Atkin et al. (2021) shows how ethnic and religious identities in India are determined by group status, group salience and the market cost of following a group's prescribed behaviors.

Benabou and Tirole (2011) introduce the idea that the choice of identity is driven by welfare maximization considerations, and that agents choose their identity to improve decision making. They develop a model in which the demand side of motivated beliefs is characterized by affective or functional benefits, such as self-esteem or self control, while the supply side is characterized by imperfect memory or awareness, which gives rise to identity investments as self signals. Agents select the way to manage memory, and consequently their identity investments, as to avoid that decisions are driven by an inappropriate rule at the interim stage. In this paper, the use of identity is similarly driven by welfare maximization considerations. The paper is therefore related to Benabou and Tirole (2011) in the sense that memory management rules in Benabou and Tirole (2011), and belief-formation rules here, are tools that agents have to improve a fitness criterion. Furthermore, these tools are useful to the extent that in their absence, agents make welfare inferior decisions; because of a systematic present bias in Benabou and Tirole (2011), and because of noisy perceptions of the chances of success here. Finally, while in both papers the mistakes agents make are determined endogenously, in Benabou and Tirole (2011) agents misperceive signals, while in this paper they see information where there is none. Furthermore, this paper shows that agents with relatively high chances of success will only use their social identity, when they belong to the socially more successful group. Pronin et al. (2004) provides evidence that is consistent with this result, and shows that female students that care much about and are strong at math actively disidentify with the character traits that they believe to be strongly related with the negative math stereotype for women.

It is not clear whether people are aware of their social identification behavior and the effects this has on their behavior. Purdie-Vaughns et al. (2008) and Marx and Goff (2005) show that black professionals and students are often aware of the presence of stereotype threat, and Steele et al. (2002) shows that some female undergraduates report in a math and science report that they believe they have weak abilities because of their gender. At the same time, Stone et al. (1999) and Leyens et al. (2000) show that white athletes and men fail to report anxiety when they experience stereotype threat. Furthermore, Banaji and Greenwald (2016) argue that the effect of social context on behavior is largely determined by the automatic part of our brain, outside of our awareness. Our brain makes automatic associations based on what we pick up from our social contexts, and often these thoughts are reflected in our actions, even when they are at times completely at odds with our conscious intentions. In the model, I make no assumptions regarding whether people are aware of their choice of belief formation process, and the model can be consistent with both scenarios.

I develop a model of endogenous belief formation, where the choice of belief formation process and social context are determined simultaneously, without assuming that social context has any direct effects on utility. Shayo (2009) also proposes a model in which social identification and social context are simultaneously determined. He defines a social identity equilibrium that is in steady state when each individual's behavior is consistent with her social identity, social identities are consistent with the social environment, and the social environment is determined by the behavior of individuals. The difference is that, in his paper, social identity directly affects preferences through built-in arguments that capture status and perceived similarity between an individual and the other members of a group. Similarly, Carvalho and Pradelski (2019) present a model in which the effect of the representation of an agent's social group on choice behavior is obtained through introducing an identity-based payoff in the utility function. Hoff and Stiglitz (2010) provide a cognitive model to explain social rigidities. They assume that people suffer from a confirmatory and preconfirmatory bias. Therefore, when groups have been historically treated as inferior, this affects how they interpret failure, which affects self-confidence. They then assume that self-confidence boosts performance as in Compte and Postlewaite (2004). In their paper, social context therefore affects utility directly through performance.

Finally, this paper is also related to the literature on evolutionary preferences, such as Dekel et al. (2007), where preferences are determined in a dynamic process according to their fitness with respect to the preferences in the rest of society. In this paper, preferences are determined according to their fitness with respect to a certain task. Hoff and Stiglitz (2016) discuss decision makers that are enculturated actors, meaning that their preferences, perception and cognition are subject to social context and cultural mental models. They discuss how these factors shape behavior through the endogenous determination of preferences and the lenses through which individuals see the world. Although their model is very different, my paper is consistent with the general ideas they bring forward. Furthermore, the model relates to the literature on stereotypes (Bordalo et al., 2016) and narratives (Akerlof and Snower, 2016), where in this paper these concepts are determined endogenously. Finally, the paper contributes to the literature on collectively sustained beliefs (Piketty, 1995; Benabou and Tirole, 2006; Frick et al., 2018), discrimination (Peski and Szentes, 2013), and affirmative action (Coate and Loury, 1993), by showing how differences in beliefs across social groups can be collectively sustained and create an equilibrium in which a priori identical individuals act differently without introducing any direct interaction between agents, nor having a common state of the world or assuming direct effects of beliefs of others on an agent's preferences.

3 The Model

3.1 The Environment

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

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

The *Non-Competitive* task has a probability of success $\gamma \in [0, 1]$ that is known and the same for all agents. Therefore, for all i,

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

More generally, γ can be interpreted as the attractiveness of the *Non-Competitive* task relative to the *Competitive* task. Agents make a series of choices during their lifetime between *Non-Competitive* tasks and *Competitive* tasks of which the probability of success depends on similar individual characteristics. For example, α_i may be related to a person's mathematical ability or leadership qualities. Although it is realistic to allow α and γ to vary slightly from choice set to choice set, the main insights of this model can also be transmitted with a simpler and more parsimonious model. Therefore, in the following, I assume that the variables α_i and γ are fixed over the lifetime of agents[9].

Noisy Perceptions - The probability α_i is unobservable, and agents only have a noisy perception $\hat{\alpha}_i$ regarding their own probability of success. I assume this perception stems from a noisy inference process that is unbiased, but always incomplete. Consequently, I pose that $\hat{\alpha}_i$ stems from a distribution g_{α_i}, such that $E(\hat{\alpha}_i) = \alpha_i$. The assumption that the noisy perception is unbiased can be challenged (See for example Mobius et al. (2014)). The objective of this assumption is nevertheless to show that a systematic bias in the individual-specific belief formation process is not the mechanism that drives the results in this model.

[9]See Section 5.2 for a discussion regarding the consequences of this simplification.

Social Context - Agents are described by an *observable characteristic* that represents for example their gender, ethnicity or social class. This characteristic is public information, meaning that both the agent herself and all other agents in the society can observe it. I let the binary[10] variable θ_i with realizations $x \in \{0, 1\}$ denote this characteristic. Then, p_x is the fraction of the population with an observable characteristic $\theta_i = x$, and each agent is fully described by her type $\{\alpha_i, \theta_i\}$. To isolate the mechanism through which social identity affects choice behavior in this model, I assume the probability α and the observable characteristic are independently distributed over the population.[11]

Agents have access to public data. This public data consists of 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. To best illustrate the mechanism that drives the results in this model, I focus on agents that pay attention to the fraction of successful individuals in the *Competitive* task with their observable characteristic among all successful individuals in this task. In the Section 4.3, I discuss how different structures on information affect behavior.

Let $\mathcal{N}_{C,x} = \{i \in N, \theta_i = x, a_i = C\}$ be the set of all individuals of type $\theta_i = x$ 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 $\mathcal{N}_{C,x} \subset \mathcal{N}_C$. Society then provides the statistic,

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

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

[10]The model can also account for non-binary observable characteristics.

[11]In Section 5, I show that the results derived from this model are generally robust in a setting where α is correlated with θ or when social identity has a direct effect on success.

[12]Agents only consider the successful individuals that have chosen the task and not all agents that have chosen the task. This captures the survivors bias (Denrell, 2003), but this bias does not drive the results.

Subjective Belief Formation - I model agents that have an imperfect idea about their economic environment and think that using their social context could be useful to form a belief about their probability of success α_i, even if they are not a priori sure of that. I therefore introduce the following family of belief formation processes with which agents form a subjective belief $\hat{p}_i$ about their probability of success of a *Competitive* task α_i, and assume agents have some discretion in finding out which belief formation process suits them best. Specifically, I have the following story in mind. Assume that agents have a natural 'urge' to look at others like them when they are not sure what to do, and people have the option to either *Repress* or *Not Repress* this urge. Let η be a 'response function' that is non-decreasing, such that

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

Furthermore, let $\eta_x = \eta(\pi_x, p_x)$. Then, agents choose a strategy $\sigma_i \in \{R, NR\}$, and

$$\hat{p}_i = \begin{cases} \hat{\alpha}_i & \text{if } \sigma_i = R \\ \eta_x \hat{\alpha}_i & \text{if } \sigma_i = NR \end{cases} \tag{4}$$

I envision agents that attempt to form a subjective belief $\hat{p}_i$, and depending on whether they let their belief-formation process be influenced by social context, their subjective belief can take two values[13] ; $\hat{p}^R$ or $\hat{p}^{NR}$. With a subjective Bayesian interpretation in mind, $\sigma_i = R$ corresponds to a world view in which private and observable characteristics are uncorrelated, while $\sigma_i = NR$, corresponds to a world view in which private and observable characteristics are correlated, with (π_x, p_x) informing about the sign and strength of that correlation[14]. Consequently, when choosing $\sigma_i = NR$, the agent biases her noisy perception $\hat{\alpha}_i$ in the direction contingent on her social type. 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 cues in the belief-formation process

[13] I do not model precisely the agent's thought processes leading to these two possible beliefs. The objective is not to propose a particular functional form, nor to root it in a specific subjective Bayesian model, but to investigate how properties of the response function can be conductive to the phenomenon I mean to describe.

[14] In this case, agents make themselves believe that the probability of success of agents like them has some predictive value for their own probability of success, and the model can be interpreted as agents exhibiting an attribution error.

leads to an optimistic interpretation of the noisy perception $\hat{\alpha}$, while this leads to a pessimistic interpretation of the noisy perception when the agent's subgroup is under-represented among the successful individuals in the society[15].

Subjective Utility Maximization - 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_i to maximize $E(u_i)$ given her subjective belief $\hat{p}_i^\sigma$, and will therefore choose the *Competitive* task if and only if $\hat{p}_i^\sigma > \gamma$. One could say therefore that agents are subjectively rational given the process that determines their subjective beliefs. Furthermore, the model allows for two different interpretations. One interpretation is that the instrument σ_i mechanically alters the agents' subjective belief $\hat{p}_i^\sigma$, where $\hat{p}_i^\sigma \in \{\hat{p}_i^R, \hat{p}_i^{NR}\}$. Another interpretation is that agents have the option to use the social identity cue to alter choice in a direction contingent on their observable type. Formally, subjective expected utility maximization implies that the agent is effectively comparing two thresholds, such that agent i chooses $a = C$ if and only if $\hat{\alpha}_i > \gamma_i$, where

$$\gamma_i = \begin{cases} \gamma & \text{when } \sigma_i = R \\ \frac{\gamma}{\eta_x} & \text{when } \sigma_i = NR \end{cases} \tag{5}$$

The use of the social identity cue in the belief-formation process implies that the agent inflates or deflates the threshold for $\hat{\alpha}$ above which she thinks she is 'good enough' to undertake the *Competitive* task. The strategy set can therefore also be directly specified as the choice set $\gamma_i \in \{\gamma, \frac{\gamma}{\eta_x}\}$. This choice set is different for agents with different observable characteristics θ, which will be the key driver of the equilibrium results.

3.2 The Solution Concept

Whether agents let social context affect their belief formation affects their choice of task. This choice behavior leads to outcomes that induce social identity cues. These social identity cues again affect the way agents form subjective beliefs. To tractably capture the fixed points of such a dynamic process, I use a static solution concept in which I assume that, given a social context, agents choose their strategy σ according to its fitness value. This fitness is determined by an individual optimality criterion I define and justify below. I then define a population equilibrium as a fixed point in social context that is induced by individually optimal strategy choices. This solution concept is in line with the view that the optimal choice of the strategy σ arises from a learning process that

[15]In the specific case where $\pi_x = p_x$, the two strategies *Repress* and *Not Repress* are equivalent.

operates faster than the dynamics in social context, where the learning of the optimal strategy happens during the lifetime of an agent through her experience with similar tasks, while changes in social context arise from agents belonging to different generations making a specific choice of task once in their lifetime.

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

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

This probability Φ is determined objectively conditional on the choice of strategy σ_i. From an outsiders perspective, the expected pay-off for agent i of type $\{\alpha, x\}$ playing σ_i given Π over all possible realization of $\hat{\alpha}$ is,

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

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

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

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

Individual optimality means that an agent uses her social identity cue to maximize her expected pay-off on average over all possible realizations of $\hat{\alpha}_i$. The fitness value of a strategy σ is therefore determined by an agent's type $\{\alpha, x\}$ and the social context Π.

I assume agents compare $V_i(R)$ and $V_i(NR)$, and choose their strategy σ_i according to the individual optimality criterion. This assumption can be justified with the view that agents have learned their optimal strategy from their own experience with similar tasks in a similar social context through for example reinforcement learning or a sampling process. The true probability α_i determines the outcomes the agent observes in this process, which enables her to learn whether it is optimal to *Repress* or *Not Repress* without precise knowledge of the relationship between her choice of strategy, choice of task and the observed outcome. Because the set of strategies is small, this is easy for agents to calculate.[16] Finally, one could say that 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_x. This aspect of bounded rationality should be considered as a modelling device that helps to keep the model parsimonious. Because of the simplifying assumption that α and γ are fixed, 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. We shall see in Section 5 how the results extend to the more realistic case in which α and γ vary or to the case in which such learning would be imperfect.

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

$$\tilde{\pi}_x(\sigma,\Pi) = \frac{p_x \int \alpha \Phi_{\alpha,x,\sigma,\Pi} f(\alpha) d\alpha}{\sum_{x\in\{0,1\}} p_x \int \alpha \Phi_{\alpha,x,\sigma,\Pi} f(\alpha) d\alpha} \tag{8}$$

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

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

In other words, a population equilibrium arises when all agents play their optimal strategy given their social context, and when these individually optimal strategy choices induce a fixed point in social context.

[16] An alternative foundation for this behavior could be in the spirit of Benabou and Tirole (2006), where there are two selves. The calm, rational self knows the true ability type, while a less calm self can be biased by all sorts of things that can happen in the moment. The first self may then decide on a belief formation rule, while the second self chooses an action at a given point in time, given its noisy perception of ability and the earlier chosen belief formation rule. The rational self ties the hands of the 'in the moment' self to limit its adverse effects on choice.

4 The Results

4.1 The Origin of Identity-Driven Choice Behavior

Definition 1 means that a strategy is individually optimal when it maximizes expected utility over all possible realizations of the noisy perception $\hat{\alpha}$. I illustrate what the behavior implied by this criterion looks like with the following example.

Example - Consider a firm in which agents make a decision whether to pursue a career in management (C) or to pursue a clerical job (NC). Furthermore, assume these agents can observe the current pool of successful managers and their gender, and that women are underrepresented in this pool. Let $\theta_i = 0$ denote being a woman and $\theta_i = 1$ being a man and assume $p_1 = p_0$. Let $\hat{p}_i^{\sigma}(\alpha, x)$ denote the subjective belief $\hat{p}_i^{\sigma}$ implied by an agent of type $\{\alpha, x\}$ playing strategy σ_i. To illustrate behavior, consider agents that have a true probability $\alpha > \gamma$, where γ represents the probability of success when choosing a clerical job. Consequently, the welfare-maximizing choice for these agents is to pursue a management career and therefore, to maximize expected utility, they should choose the strategy σ that most increases the likelihood to choose this career. It is therefore only optimal to *Repress* the urge to look at others when $P(\hat{p}_i^{NR}(\alpha, x) > \gamma) \leq P(\hat{p}_i^{R}(\alpha, x) > \gamma)$.

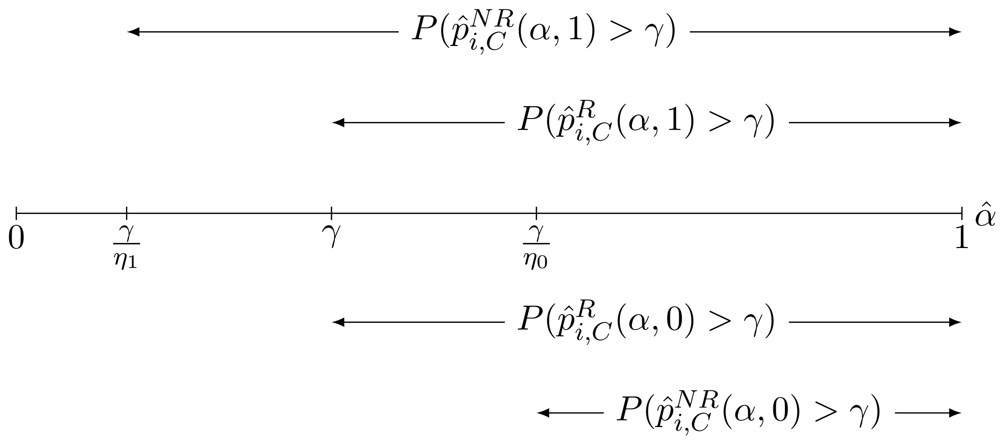

Figure 1: An Illustration of the different probabilities to choose $a = C$ for an agent with $\alpha > \gamma$ in a social context such that $\pi_0 < \pi_1$.

Figure (1) illustrates the implications of the strategies *Repress* and *Not Repress.* Because men are overrepresented among the current successful managers in the firm, *Not Repressing* the use of the social identity cue in belief formation causes men to deflate the threshold above which they think they are 'good enough' to become a successful manager. Consequently, choosing a management career becomes relatively more attractive than choosing a clerical job. Individual optimality implies therefore that men with $\alpha > \gamma$ should *Not Repress* the urge to look at others like them. For women the story is different. Because women are underrepresented among the successful managers, *Not Repressing* the use of the social identity cue in the formation of their subjective belief would inflate the threshold above which they think they are 'good enough'. This would make a management career relatively less attractive and individual optimality implies therefore that women should *Repress* the urge to look at the outcomes of other women. The opposite reasoning applies for agents with $\alpha < \gamma$. In general, one can show that agents with a relatively high true probability of success α will want to 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 probability of success α. Proposition 1 formalizes this result.

PROPOSITION 1 (Individually Optimal Belief Formation): *The individually optimal strategies σ^* given an agent's type $\{\alpha, x\}$ are the following:*

- *The individually optimal strategy σ^* is 'Not Repress' for agents of type $\{\alpha, x\}$ such that $\alpha > \gamma$ and $\pi_x > p_x$ or $\alpha < \gamma$ and $\pi_x < p_x$*
- *The individually optimal strategy σ^* is 'Repress' for agents of type $\{\alpha, x\}$ such that $\alpha > \gamma$ and $\pi_x < p_x$ or $\alpha < \gamma$ and $\pi_x > p_x$*

The use of social identity cues gives agents the option to nudge beliefs in a particular direction determined by the social context. Proposition 1 demonstrates that, when beliefs are biased towards the welfare-maximizing task, this enables agents to improve decision making on average by increasing the likelihood that they choose the welfare-maximizing task over all possible realizations of $\hat{\alpha}$.[17] This shows how choice behavior can be driven by observable characteristics and social context, even when these observable characteristics and social context have no direct effect on utility.

[17]One might argue that, if agents are able to learn their optimal strategy σ conditional on α, they should also be able to retrieve their true value of α from this optimal strategy. I believe this way of thinking is too much driven by the simplification of the model in which α and γ are fixed over the lifetime of an agent, and I will elaborate more in the Discussion section on how the model can account for sophisticated agents that understand what their fitness signals about their true probability of success.

Talent will always find its way - The ability to improve decision making on average using social identity cues is a function of the true probability α. Specifically, if we assume the variance of $\hat{\alpha}$ is uncorrelated with α, the model predicts that the behavior of agents with true probabilities of success close to γ will on average be influenced more by 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 choosing *Repress* is given by $P(\hat{\alpha} < \gamma)$. Because $\hat{\alpha}$ is unbiased, the further away the true probability α is from γ, the lower the chances that the agent has a belief $\hat{\alpha} < \gamma$[18]. A similar reasoning applies to agents with $\alpha < \gamma$. Therefore, the use of social identity cues in belief formation is on average most beneficial to those who have a true probability of success α close to γ, while agents with extremely low or extremely high values of α 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 - Proposition 1 implies that, when agents use their social identity cue in belief formation, this leads to an either optimistic or pessimistic processing of their noisy perception $\hat{\alpha}_i$ that is always welfare improving. This optimism or pessimism gives agents therefore the option to manage their over- or under-confidence regarding their chances of success, as captured in their perception $\hat{\alpha}_i$. The use of social identity cues can be interpreted as an instrument that gives the agent the option to manage confidence, and Proposition 1 shows that the optimal use of social identity cues results in the optimal management of confidence to improve decision making on average. The model therefore sheds light on the instrumental value of both over- and under-confidence to improve decision making.[19] 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 can be asymmetric across agents.

[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]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. (Mobius et al., 2014) describe how this applies to agents who use optimistic processing that leads to over-confidence. 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 intact.

4.2 The Persistence of Identity-Driven Choice Behavior

When agents choose σ_i^* in a social context where one subgroup is overrepresented among the successful individuals, this affects both how many and what type of agents choose the *Competitive* task. Specifically, when $\theta_i = x$ implies a more pessimistic processing of the noisy perception $\hat{\alpha}_i$, then those who choose the *Competitive* task despite this, tend to have a larger success rate on average than those who choose *Competitive* task with a characteristic implying an optimistic processing of $\hat{\alpha}_i$. This is what we call the '*selection effect*'. On the other hand, the population of those that belong to the socially less successful subgroup choosing the *Competitive* task tends to be smaller than the population of those choosing the *Competitive* task belonging to the socially more successful subgroup. This is what we call the '*population effect*'. Corollary 1 formalizes this.

COROLLARY 1: *Let* $x' \in \{0,1\}$ *be the complement of* x *and assume WLOG that* $\pi_x > \pi_{x'}$. *The optimal use of social identity has both a population effect, such that* $\Phi_{\alpha,x,\sigma_i,\Pi} > \Phi_{\alpha,x',\sigma_i,\Pi}$ *and a selection effect, such that* $E(\alpha|a = C, x) < E(\alpha|a = C, x')$. *The strength of both effects is such that the order* $\pi_x > \pi_{x'}$ *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. When we assume agents choose their strategy σ_i according to the *Individual Optimality* criterion, Proposition 1 implies that men with $\alpha > \gamma$ will choose *Not Repress*, while women with $\alpha > \gamma$ will choose *Repress*. Similarly, women with $\alpha < \gamma$ will choose *Not Repress*, while men with $\alpha < \gamma$ will choose *Repress*.

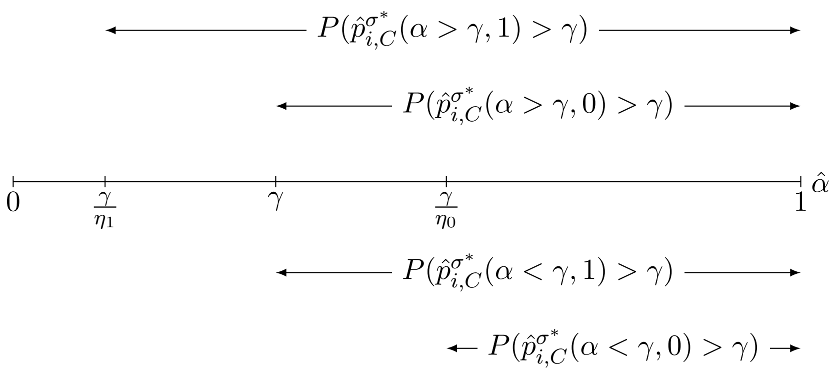

Figure 2: Optimal Decision-Making in a social context such that $\pi_0 < \pi_1$. The part above the axis describes for which realizations of $\hat{\alpha}$ agents correctly choose $a = C$, while the part under the axis shows for which realizations of $\hat{\alpha}$ agents incorrectly choose $a = C$.

Figure (2) shows the realizations of $\hat{\alpha}$ for which men and women choose to pursue a career in management or a clerical job. Because the number of agents in the society is arbitrarily large, the probabilities depicted in the figure 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 $\hat{\alpha} > \frac{\gamma}{\eta_1}$. Because $\frac{\gamma}{\eta_1} < \gamma$, 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 $\hat{\alpha} > \frac{\gamma}{\eta_0}$. Since $\gamma < \frac{\gamma}{\eta_0}$, 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 mistakenly choose a clerical job when they receive a realization $\hat{\alpha} < \frac{\gamma}{\eta_1}$. Again, Figure (2) shows that the latter event is less likely. We therefore conclude that more men will, both optimally and suboptimally, choose to pursue a management career, while more women will, both optimally and suboptimally, choose a clerical job. This demonstrates the '*population effect*'. Finally, because the noisy perception is unbiased, higher realizations of $\hat{\alpha}$ are more likely for agents with higher true probabilities α. Because $\gamma > \frac{\gamma}{\eta_1}$, men choose the management task for on average lower realizations of $\hat{\alpha}$ than women and therefore, $E(\alpha|a = C, 1) < E(\alpha|a = C, 0)$. In other words, the average success rate of men in a management career will be lower than the average success rate of women. This demonstrates the '*selection effect*' [20]. The intuition behind the result that the selection and population effect will not reverse the order $\pi_1 > \pi_0$, is the following. Because α and θ are independently distributed, the number of men with $\hat{\alpha} > \gamma$ in an arbitrary large population is equal to the number of women with $\hat{\alpha} > \gamma$. The population of men that optimally choose a management career consists therefore of the men that have a belief $\hat{\alpha} > \gamma$ plus the more able men that have a belief such that $\frac{\gamma}{\eta_1} < \hat{\alpha} < \gamma$, while the population of women that optimally choose a management career only consists of the more able women that have a belief $\hat{\alpha} > \gamma$. Therefore, even though the women that choose a management career have on average a higher success rate, the expected number of successful men is always larger than the expected number of successful women.

[20] 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.

Existence and Stability of Population Equilibria - Whether the identity-driven choice behavior induced by agents choosing their individually optimal strategies can be persistent, depends on whether the population and selection effect reinforce or shrink differences between π_x and $\pi_{x'}$. 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, identity-driven choice behavior could also induce persistent asymmetric allocations over tasks across a priori identical subgroups. The two different scenarios are defined in Definition 3.

DEFINITION 3 (Equilibrium Regimes): *In a '**Neutral Regime**' the allocation of individuals over the different tasks is symmetric across different subgroups, and $\pi_x = p_x$. In a '**Non-Neutral Regime**' the allocation of individuals over the different tasks is asymmetric across different subgroups, and $\pi_x \neq p_x$.*

A '*Neutral Regime*' always exists. The intuition behind this result is the following. Assume that we are in a social context such that $\pi_x = p_x$. In this case, the strategies *Repress* and *Not Repress* are equivalent. Therefore, no matter what strategy σ agents choose, there will be no differential influence of the social context on belief formation across different social groups, and $\tilde{\pi_x}(\sigma, \pi_x) = \pi_x$. Consequently, I use a stability criterion to analyze when a set of individually optimal strategies and a social context constitute together a stable configuration. Let $\rho = \frac{\pi_x}{\pi_{x'}}$ and $\tilde{\rho} = \frac{\tilde{\pi}_x(\sigma,\pi_x)}{\tilde{\pi}_{x'}(\sigma,\pi_x)}$. Then the stability criterion is defined as follows.

DEFINITION 4 (Stability): *An equilibrium regime is unstable, when,*

$$\frac{\partial \tilde{\rho}}{\partial \rho} > 1 \tag{10}$$

In the following example, I illustrate how a '*Neutral Regime*' becomes unstable, and how the social identity cue $\tilde{\pi}_x(\sigma, \Pi)$ induced by a social context Π is bounded from above.

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 response function,

$$\eta(\pi_x, p_x) = \begin{cases} +\infty & \text{if } \pi_x > p_x \\ 1 & \text{if } \pi_x = p_x \\ -\infty & \text{if } \pi_x < p_x \end{cases} \tag{11}$$

When $\pi_0 = \pi_1 = \frac{1}{2}$, the strategies *Repress* and *Not Repress* are equivalent, and this social context induces a social context such that $\tilde{\pi}_1(\sigma^*, \frac{1}{2}) = \frac{1}{2}$. In other words, a '*Neutral Regime*' exists. Nevertheless, as soon as agents observe slightly more men than women among the successful managers, such that $\pi_1 > \pi_0$, then, under the assumption that agents choose σ according to the individual optimality criterion, the extreme response function $\eta(\pi_x, p_x)$ causes all men with $\alpha > \gamma$ to choose to pursue a management career, while all women with $\alpha < \gamma$ will choose the clerical job. Consequently, $\tilde{\pi}_1(\sigma^*, \pi_1) > \pi_1$, while $\tilde{\pi}_0(\sigma^*, \pi_0) < \pi_0$ and the 'Neutral Regime' becomes unstable. In this extreme case, we can show that the induced social identity cue $\tilde{\pi}_1(\sigma, \Pi)$ is bounded from above. Specifically, let $S_1 = \lim_{N\to\infty} \frac{N_1}{N}$ be the fraction of successful male managers. The upper bound $\overline{S}_1$ on S_1 is characterized by the fact that 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,

$$\overline{S}_1 = p_1 \int_{\alpha<\gamma} \int_{\hat{\alpha}>\gamma} \alpha g_\alpha(\hat{\alpha}) f(\alpha) d\hat{\alpha} d\alpha + p_1 \int_{\alpha>\gamma} \alpha f(\alpha) d\alpha \tag{12}$$

Similarly, let $S_0 = \lim_{N\to\infty} \frac{N_0}{N}$ be the fraction of successful female managers. The lower bound $\underline{S}_0$ on S_0 is characterized by the fact that only women with $\alpha > \gamma$ and $\hat{\alpha} > \gamma$ and no women with $\alpha < \gamma$ will choose to pursue a management career. Consequently,

$$\underline{S}_0 = p_0 \int_{\alpha>\gamma} \int_{\hat{\alpha}>\gamma} \alpha g_\alpha(\hat{\alpha}) f(\alpha) d\hat{\alpha} d\alpha \tag{13}$$

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

$$\overline{\pi}_1(\sigma, \Pi) = \frac{\overline{S}_1}{\overline{S}_1 + \underline{S}_0} \tag{14}$$

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

Because $\overline{\pi}_\theta(\sigma, \Pi)$ is bounded from above, I derive a sufficient condition for the existence of a stable '*Non-Neutral Regime*' by analyzing when a '*Neutral Regime*' becomes unstable when we perturb it slightly. This condition is presented in Proposition 2.

PROPOSITION 2 (Existence Non-Neutral Regime): *Let* $p_0 = p_1$*, and let* $\delta > 0$ *be a small value with which we disturb a 'Neutral Regime'. A sufficient condition for the co-existence of a stable 'Non Neutral Regime' with a 'Neutral Regime' is as follows,*

$$\frac{\partial \gamma^{NR}}{\partial \delta} \frac{|\frac{\partial S}{\partial \gamma}|}{S} > 4 \tag{15}$$

where $\gamma^{NR} = \frac{\gamma}{\eta_x}$ *and* $S = \int \alpha G_\alpha(\gamma) f(\alpha) d\alpha$*.*

Proposition 2 shows there are two ingredients that contribute to making a '*Neutral Regime*' unstable. First of all, a change of δ in Π away from a 'Neutral Regime' must have a sufficiently large effect on the choice behavior of agents at the individual level. Specifically, the induced change in the threshold γ^{NR} of agents that choose to 'Not Repress' the use of the social identity cues must be large enough. This change depends first of all on the derivative of the response function $\eta(\pi_x, p_x)$ at the '*Neutral regime*'. Furthermore, because of the linearity of γ^{NR} in γ, this change is multiplicative in γ. The second ingredient that contributes to the instability of a '*Neutral Regime*' is driven by the effect of a change δ on the outcomes at the aggregate level. This is captured by the elasticity of S in γ, where S is the total number of successful people at the *Competitive* task. The absolute value of this elasticity is increasing in γ, since the more attractive the outside option, the lower the number of agents S that tries and succeeds at the *Competitive* task. Furthermore, the higher γ, the higher the α of agents that choose the task and do 'Not Repress' the use of social identity cues. The intuition behind the positive effect of this elasticity $\frac{|\frac{\partial S}{\partial \gamma}|}{S}$ on the instability of a '*Neutral Regime*' is therefore as follows. The higher γ, the higher the expected success rate of those who change their choice behavior due to small changes in the social context away from the '*Neutral Regime*'. Furthermore, the higher γ, the smaller the total number of successful people and the larger the effect of a change in behavior of a small group of agents on the induced social identity cues $\tilde{\Pi}(\sigma, \Pi)$. Whether a '*Non-Neutral*' is unstable depends therefore on γ and the properties of the response function $\eta(\pi_x, p_x)$.

Minority Effect - There are two different ways in which agents can process π_x and p_x in their response function. They can either process the difference $\pi_x - p_x$ or the proportion $\frac{\pi_x}{p_x}$. When $p_0 = p_1$, the local effect of a small change in a 'Neutral Regime' for a response function in which agents process the difference or the proportion are the same. This is nevertheless not true when $p_0 \neq p_1$. The effects of a change away from the '*Neutral Regime*' are symmetric for both social groups when agents process the difference $\pi_x - p_x$, independent of the values of p_x, and therefore, even when $p_0 \neq p_1$, Proposition 2 is still correct. When agents process the proportion $\frac{\pi_x}{p_x}$ in their response functions, the effects of a change in the social context on $\tilde{\pi}_x$ and $\tilde{\pi}_{x'}$ are not symmetric anymore. Specifically, with such a response function, the minority group will react more strongly to changes in the social context than the majority group, which is also reflected in the sufficient condition that can be derived for this case.

Degree of Asymmetry - The more strongly agents react to their social context in their belief formation process, the more the social context will drive their choice behavior. Corollary 2 shows how this affects the asymmetry we observe in a '*Non-Neutral Regime*'.

COROLLARY 2: *Take two response functions $\hat{\eta}$ and η, such that $\hat{\eta}(\pi_x, p_x) > \eta(\pi_x, p_x)$ for all $\pi_x > p_x$. 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 response function η. Then, $\pi^*_{\hat{\eta},1} > \pi^*_{\eta,1}$.*

Corollary 2 shows that the function η determines the degree of asymmetry we observe across the different social groups in a 'Non-Neutral Regime', where the difference between π_x and p_x is larger, the stronger agents let their subjective beliefs be influenced by their social identity cues. This also implies that, when the minority group reacts more strongly to changes in the social context than the majority group, their social identity cue in equilibrium will be further away from the '*Neutral Regime*' than the social identity cue of the majority group.

Welfare - If we consider a social planner that aims to maximize the aggregate utility of all agents in the society, then a 'Non-Neutral Regime' creates a Pareto improvement with respect to a 'Neutral Regime'. In a 'Neutral Regime', the strategies *Repress* and *Not Repress* are equivalent, and all agents form beliefs in the same way. In a 'Non-Neutral Regime', only those agents for whom it is strictly optimal will take the asymmetries in the social context into account. Those agents, for whom it is no improvement to let their social context influence their belief formation, will not change their behavior. Therefore, a 'Non-Neutral Regime' makes the agents that can use their social identity cue to improve decision making better off compared to a 'Neutral Regime', but the agents that cannot use their social identity cues to improve decision making will not be made worse off by a 'Non-Neutral Regime'. The Pareto optimality of a 'Non-Neutral Regime' should nevertheless be considered as a result that is driven by the simplifying assumptions made in the model, and can be easily contested. For example, a 'Non-Neutral Regime' is not necessarily Pareto optimal when beliefs have a direct effect on the probability of success, through for example confidence (Compte and Postlewaite, 2004), or when social context has a direct effect on someone's chances of success through some form of discrimination or stereotype threat (Steele, 2010). Finally, a 'Non-Neutral Regime' can become suboptimal when agents make systematic errors in learning their optimal belief formation process, when they do not correctly compute the long-term pay-offs of choosing a *Competitive* task, or when the strategy *Repress* becomes costly.

4.3 Social Identity Cues: Structure and Information

In this section, I provide insights into how different structures on the social identity cues affect the equilibrium results. When introducing the social identity cues, I assumed that people could not observe those who tried but failed. Although this assumption may be justified in a variety of settings, there are also settings in which this assumption may not hold. The key aspect is that what matters is not whether agents can observe those who tried but failed, but how they use this information. There are two different ways in which agents can take this information into account. First of all, they can put a similar structure on the public information as in the in exposition of the model, but change the denominator of the social identity cue to consider all agents that have tried the *Competitive* task, instead of only considering agents that have tried and succeeded. It can be shown that this structure on the social identity cues does not eliminate the existence of a '*Non-Neutral Regime*'. On the other hand, agents may also use the information of the number of agents that tried but failed to calculate the within-group average success rate. Specifically, let S_x be the number of agents with characteristic $\theta_i = x$ that has tried the *Competitive* task and succeeded, while F_x is the number of agents that has tried but failed. Then,

$$\pi_x = \frac{S_x}{S_x + F_x} \tag{16}$$

With this different structure on π_x, the individually optimal strategies are determined similarly by an agent's type $\{\alpha, x\}$ and by whether $\pi_x > \pi_{x'}$ or vice versa, and, at the aggregate level, choice behavior would look as depicted in Figure (2). The effect of the aggregate choice behavior on the induced social identity cue $\tilde{\pi}_x(\sigma, \pi_x)$ will nevertheless be very different. To illustrate, consider an example in which the average success rate of men, π_1, is higher than the success rate of women, π_0. Figure (2) shows that such a social context induces more men to choose the *Competitive* task than women, but that this induces simultaneously a higher average success rate for women than for men. Consequently, the new social identity cues will induce more women to choose the *Competitive* task than men, which will induce a higher average success rate for men, and so forth. Therefore, asymmetries in choice behavior induced by social identity cues reflecting the within-group average success rates cannot be persistent.

Finally, agents could also process statistics that only include those who try the *Competitive* task, because it may take too long for a 'successful outcome' to be realized, or it may not be universally clear what a 'successful outcome' looks like, which drives agents to look for more objective statistics. In this case, the story behind why agents believe the social identity cues are relevant could be more of an imitation story. For example, they might believe that the fraction of people that try the *Competitive* task among those in their social group is indicative for their suitability to the task, or the degree of hostility in the environment (Chung, 2000). Although various statistics reflecting those who try could be considered, all of these statistics allow the existence of a '*Non-Neutral Regime*'.

In general, this discussion sheds light on what structures on information drive the persistence of identity-driven choice behavior in this model. The key aspect that causes asymmetry in the social identity cues is the population effect induced by identity-driven choice behavior. To enable the existence of a '*Non-Neutral Regime*', the structure agents put on their public information must be able to reflect this effect. The within-group average success rates only reflect the selection effect, and therefore, when people focus on this statistic, a '*Non-Neutral Regime*' cannot exist. In all other cases discussed, the induced social identity cues are driven by the population effect, and therefore a '*Non-Neutral Regime*' can arise.

5 Discussion

5.1 Imperfect Learning

The assumption that agents choose their individually optimal strategies can be justified with the view that agents learn these strategies through their own experience with similar tasks. In the previous sections, I assume that agents are perfectly able to learn these individually optimal strategies. The main objective of this assumption is to provide a benchmark and show that, even when agents are able to perfectly learn how to optimally use the social identity cues, asymmetries in choice behavior can still persist in equilibrium. It is nevertheless likely that agents are only able to learn imperfectly. In this section, I discuss that, as long as the probability to learn something is positive, a '*Non-Neutral Regime*' can still exist.

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 (6) can be written as,

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

In the case of perfect learning, $P(\sigma_i = \sigma|\alpha, x, \Pi) \in \{0, 1\}$, while in the case of imperfect learning, $P(\sigma_i = \sigma|\alpha, x, \Pi) \in [0, 1]$. Let λ be an exogenous learning process. Then, any such learning process implies a probability $P^{\lambda}(\sigma_i = \sigma|\alpha, x, \Pi)$. For example, let $V(\sigma)$ be the expected pay-off of playing strategy σ and let $\lambda = QR$ represent a quantile response model. Then, $P^{QR}(\sigma_i = \sigma|\alpha, x, \Pi) = \frac{exp(\beta V(\sigma))}{\Sigma_{\sigma \in \{R,NR\}} exp(\beta V(\sigma))}$, 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, $P^S(\sigma_i = \sigma|\alpha, x, \Pi) = P(\hat{v}(\sigma) > \hat{v}(\sigma'))$, 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.

As long as the probability to learn is strictly positive, there will be a discrepancy between $\Phi_{\alpha,x,\sigma_i,\Pi}$ 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 from experience. When the variance of $\hat{\alpha}$ is uncorrelated with α, this is on average more likely to happen to agents with ability levels close to γ. As discussed earlier, these agents also happen to be the agents whose decision making will be most affected by the social context. 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 will most likely 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 influence of the social context when this is optimal, imperfect learning implies that more agents than optimal use their social identity cue in belief formation. In this case, imperfect learning would increase the strength of the population effect. If failing to learn implies that agents make random mistakes, imperfect learning decreases the strength of the population effect. How people learn is eventually an empirical question, and I talk more about how this issue provides interesting directions for future research in the Conclusion.

5.2 Inference based on the Individually Optimal Strategy

To simplify the exposition of the model, I assume that the true probability α and the probability of success of the outside option γ are fixed over an agent's lifetime. Although this assumption allows me to show the mechanism through which the social context affects decision making in the simplest way, it may also raise the question why, if agents are able to learn their individually optimal strategy conditional on their true value of α, they are not able to retrieve their true probability of success from this information. Indeed, a sufficiently sophisticated agent could interpret her individually optimal strategy as an extra signal regarding her true probability of success. In the simplified version of the model, this signal would be a perfect signal regarding the true α, and, because α and γ are fixed, it becomes optimal to either always choose the *Competitive* task or always choose the *Non-Competitive* task. Therefore, agents that are sophisticated enough to know the structure of the model could be able to find out which choice of task is optimal from learning their optimal strategy. This reasoning is nevertheless too much driven by the simplifying assumptions in the exposition of the model.

In a more realistic model, the values of α and γ vary over the lifetime of an agent, and agents learn from a series of related, but slightly different tasks. For example, an agent learns the strategy σ^* used to make the decision whether to choose a STEM major from earlier experiences with STEM related tasks throughout high school. The individually optimal strategy is in this case conditional on whether on average during the learning process α has been above or below γ. When modelling the choice to enter a STEM major, we will therefore have a fraction of agents with $\alpha > \gamma$ that belong to the socially more successful group that will choose *Not Repress*, but also a fraction that will play *Repress*.

Consequently, the optimal strategy becomes an imperfect signal regarding α. Sophisticated agents will therefore no longer be able to retrieve their true α from this signal. They may still want to use this extra signal to improve upon their decision, which eventually means forging a third belief: $\hat{p}_i^I$, the resulting belief from the inference process I[21]. This sophisticated agent, who thinks further about the consequence of finding out wheter *Repress* is a good strategy and is willing to test $\hat{p}_i^I$, is just an agent that ends up comparing three possible beliefs; $\hat{p}_i^R, \hat{p}_i^{NR}$ or $\hat{p}_i^I$. Without further information to exploit, she will not be able to further improve decision making[22]. The model can therefore account for this type of sophistication, which simply amounts to enriching the set of strategies that is considered[23].

To analyze the effects of such an enriched set of strategies on the equilibrium results, consider again the example of the firm. If the belief $\hat{p}_i^I$ results from a correct inference process, then a man that learns that *Not Repress* is optimal for him, will be able to make the inference that, on average, he is good enough to choose leadership-related tasks, while a man that learns that *Repress* is optimal, will make the inference that, on average, he is not good enough to choose leadership-related tasks. The opposite applies to women. In this case, the key aspect that causes a 'Non-Neutral Regime' to exist, namely the fact that the belief-formation process is type-contingent, disappears. One could argue nevertheless that the latter inference process seems much more difficult to make than the former, since it requires a more elaborate thinking. It seems easier for a man to realize that, when he should use the fact that males are overrepresented among the successful individuals, then he is probably also likely to succeed. But it is much less straightforward for a man to conclude that, if it is not optimal to use the male-driven bias, then it must be that he has low chances of success. He may simply interpret this information as the statistic not providing useful or relevant information regarding his own abilities. Similarly, for a female, if she learns *Not Repress* is a better strategy than *Repress*, it may be easy for her to conclude that, like all other women, her chances of success are not that great either. It is much less straightforward for her to conclude that, if *Repress* is the better strategy, then she must be good. Therefore, if the belief $\hat{p}_i^I$ follows from a correct and complete inference process, a 'Non-Neutral Regime' can no

[21]This inference process does not have to be accurate nor based on the correct model for the world.

[22]The same reasoning applies to agents that can only imperfectly learn their individually optimal strategy in the simplified model.

[23]This would limit the degree to which agents can compare these strategies, since one cannot compare more strategies without at the same time altering the accuracy with which one can compare them.

longer exist. On the other hand, a partial inference process, like the one described above, would exacerbate the phenomenon, as in this case all high probability male choose the *Competitive* task, along with all low probability males that have a good draw, while all low probability women choose the *Non-Competitive* task, along with all high probability women that have a bad draw. Again, this discussion emphasizes the key aspect of the model, namely that the 'Non-Neutrality' appears when the belief-formation rules or heuristics are built on type-contingent statistics.

5.3 A Misspecified Response Function

In the exposition of the model, I assume agents hold correct beliefs about the fraction p_x of the population that has an observable characteristic $\theta_i = x$. It is not always obvious that this is the case. Corollary 3 shows what happens when these beliefs are incorrect.

COROLLARY 3: *Assume WLOG that agents hold a belief $\hat{p}_x > p_x$. Then there only exists a 'Non-Neutral Regime' in which $\pi_x < p_x$. A 'Neutral Regime' no longer exists.*

Example - Corollary 3 can be illustrated as follows. 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 the belief be $\hat{p}_0 = \frac{1}{2}$, while the true fraction is $p_0 < \frac{1}{2}$. 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. Consequently, 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. Individual optimality then induces women with $\alpha > \gamma$ to *Repress* the influence of the social context on their beliefs, while men with $\alpha > \gamma$ will not. This drives the population towards a '*Non-Neutral Regime*' in which women are indeed underrepresented among the successful managers.

5.4 Correlated Chances of Success and Observable Characteristics

To isolate the effect of the optimal use of social identity on decision making, I assume the probability α and the observable characteristic θ are independently distributed over the population. In practice, a high probability of success often results from the degree to which a certain 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) chances of success and observable characteristics[24]. Such a correlation can also be due to nature, such as muscular strength and gender (Alesina et al., 2013). This type of correlation will not affect the potential gains of using the social context in belief formation at the individual level, and will therefore not affect the individually optimal strategies as presented in Proposition 1. It will nevertheless affect the outcomes observed at the aggregate level. In most empirical examples, having high chances of success 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, which increases the strength of the population effect. Therefore, the use of social identity cues in belief formation would exacerbate 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. Although less likely, the correlation could also work in the opposite direction, where having low chances of success is correlated with belonging to the socially more successful subgroup. In this case, the correlation decreases the strength of the population effect, and, at the aggregate level, the effect of the individually optimal strategies and the effect of correlated chances of success and observable characteristics work in opposite directions.

5.5 Direct Effects of the Social Context on Success and Utility

Social identity also has direct effects on utility. Agents can derive for example utility from the acceptance by 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), and the chances of success can be affected by statistical discrimination based on the social context (Coate and Loury, 1993) or by differential network effects (Lalanne and Seabright, 2011). This implies that the probability of success α is not only a function of individual-specific characteristics, but also of the observable characteristic θ. Consequently, the social context contains real information and the agent's 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 individually optimal strategy will affect the true probability of success

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

in a way that will reinforce the population and selection effect. Therefore, in this model, any direct effects of the social context on utility and performance will be complementary to the effects of the individually optimal strategies, and both effects will be simultaneously and endogenously determined through the social context Π. The results presented in Propositions 1 and 2 are therefore robust in this context, and any direct effects of the social context on utility and performance increase the strength of the population effect.

5.6 Correlation Noise in Beliefs and True Probabilities of Success

One could argue that the individual beliefs $\hat{\alpha}$ could be noisier for agents with low true chances of success than for agents with high chances of success, because agents with high probabilities α are more likely to undertake similar tasks and will therefore receive more feedback regarding their underlying abilities. The noise in $\hat{\alpha}$ could also be correlated with the observable characteristic, because agents with an observable characteristic that makes them less likely to choose a *Competitive* task conditional on their value of α are also less likely to receive feedback about their underlying abilities. Finally, one could pose that agents with extreme true probabilities of success have beliefs $\hat{\alpha}$ that are less noisy than agents with average values of α.

Agents with beliefs $\hat{\alpha}$ that are more noisy can potentially gain more from the optimal use of the social context, because they are more likely to make mistakes. Those agents that can indeed use their social context to improve decision making can therefore compensate this higher degree of noisiness with an optimal management of confidence. Agents that cannot use their social context to improve decision making will not have an instrument to off-set their higher likelihood to make mistakes. Because the social context drives choice behavior more when there is more noise in $\hat{\alpha}$, the effect of all above-mentioned correlations between the noise in $\hat{\alpha}$ and α will strengthen the population effect.

6 Applications and Possible Tests of the Model

6.1 Perceptions and Role Models

The model predicts that, when people are uncertain about their chances of success, their choice behavior can be driven by what they perceive in their social context. Moreover, depending on how people process the information stemming from their social context, these perceptions can become self-fulfilling in equilibrium or not. This result has interesting policy implications. It implies that people can be nudged towards desirable behavior by influencing the statistics that people take into account, or by making hidden data, such as those who tried but failed, more visible. This can be a policy complementary to a real and maybe more costly change of the social context, through for example affirmative action policy. The model also 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 can work through the same mechanism as in Chung (2000), where agents use role models to retrieve information about the degree of hostility of the environment they will work in. Another way to interpret the effect of role models on decision making in this model is that agents believe they can retrieve information regarding their individual-specific chances of success from people like them.

6.2 Individual Feedback

Because the options to manage confidence using the social context are asymmetric across social types, similar types of individual feedback can have different effects on choice behavior for agents belonging to different social groups. One could exploit these differences to create more diversity in educational and professional environments. Specifically, the effects of a 'Non-Neutral' social context could be offset by creating a similar bias in the noisy belief for exactly those agents that cannot optimally bias their decision making using 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 perception upwards in a similar way as what men can achieve with the use of the social context. These insights can be useful for the development of programs that aim to enhance the confidence of underrepresented groups to increase diversity.

6.3 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 in belief-formation, and the incompleteness of the learning process of the chances of success. The asymmetries in choice behavior across different social groups that are induced by belief formation that is contingent on social types, 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. 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. Where the current literature predicts that the only way to effectively fight harmful stereotypes is to change the reality in which people make decisions, this model provides an extra insight, namely that nudging people towards the use of desirable statistics or a desirable way to process the information in the social context could also be effective.

6.4 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 response function of agents, by their sensitivity to changes in the social context, and by their beliefs p_x about 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 as being underrepresented, then, as soon as the policy is removed, choice behavior driven by type-contingent belief-formation processes will automatically bring the society back to a 'Non-Neutral Regime'. Such temporary affirmative action will therefore not have any long-run effects.

Furthermore, when agents are perfectly able to learn their optimal strategies, the effects of a quota in this model are probably not achieved in the way policy makers would desire. Specifically, the changes in the representation of social groups are obtained through a loss of agents belonging to the socially more successful group with $\alpha > \gamma$ that choose the *Competitive* task, and a gain in agents belonging to the socially less successful group with $\alpha < \gamma$ that choose the *Competitive* task. This sheds light on how affirmative action policy can potentially hurt the decision making of both agents, which decreases aggregate welfare.

6.5 Oppositional Identities

The phenomenon of oppositional identities describes 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'. The model shows that, even in the absence of social pressures or cultural transmission, we could observe a disproportionate part of the minority students reject the dominant culture when these students observe that students like them are under-represented among the successful students that work hard in school. The model argues therefore that minorities adopting 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 in this model driven by exhibiting a lack of confidence regarding one's own chances of success and not having the option to offset this lack of confidence using information from the social context.

7 Conclusion

This paper analyzes the origin and persistence of identity-driven choice behavior using an equilibrium model of endogenous subjective belief formation in social context. To shed light on the origin of identity-driven choice behavior, the model shows how social identity cues that are informationally irrelevant in decision making can be used to optimally bias noisy perceptions to limit the adverse effects of these perceptions being incorrect. 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 there exist correlations between success and certain observable characteristics. The model shows that, if we manage to attribute success or failure to certain observable characteristics in an optimal way, then this would enable us to form beliefs that improve our decision making on average.

The options available to bias beliefs depend on social context, and can be asymmetric across agents with different observable characteristics. At the aggregate level, this asymmetry can create 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. Therefore, even when agents manage to behave optimally from an individual perspective, asymmetries in choice behavior across a priori identical subgroups can persist.

The results imply that choice behavior driven by observable characteristics is not necessarily caused by the explicit preferences of agents with these observable characteristics, but could driven by the social context in which these agents make their decisions. This suggests that behavior following from social norms or stereotypes, as described in Akerlof and Kranton (2000), does not have to be driven by the internal manifestations of an identity, but can rather be 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 belief formation that are induced by the social context, biases that may help people at the individual level, but are not necessarily beneficial for society. This paper provides insights that can help to develop the adequate informational policies that could limit the harmful effects of these biases at the aggregate level.

Finally, the insights provided in this paper point to several areas for future research to better understand the impact of identity-driven choice behavior in our society. First, 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 social context across agents. Moreover, this heterogeneity in perceptions could be correlated with observable characteristics through variables such as income, neighborhood or education. An interesting way in which this model could therefore be extended is by introducing heterogeneity and network effects in the perceptions of the social identity cues, and analyze what the effect would be on the equilibrium results.

Secondly, the results in this paper are derived under the assumption that agents are perfectly able to learn their optimal belief formation strategies. The goal of this assumption was to simplify the exposition of the behavioral phenomenon I introduce, and to create a benchmark for the equilibrium results. The process in which agents learn how to interpret and use social context is nevertheless an interesting research topic on its own. For example, social context could influence this learning process directly, through for example stigmatization, discrimination, implicit biases and expectations, social pressures or stereotype threat, which could induce learning traps that could be asymmetric across social groups. Psychological factors may also play a role, such as the shame to learn you are not good enough to undertake a task, even though you belong to socially more successful group. This could affect the set of belief formation heuristics people choose from. To better understand how identity-driven choice behavior plays a role in the inequalities we observe in our society, we need to obtain a deeper understanding of this learning process with the use of both theoretical and empirical research. This would allow us to better make the step from the theoretical framework to the real world and derive more concrete policy implications.

References

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

G.A. Akerlof and D.J. Snower. Bread and bullets. *Journal of Economic Behavior and Organization*, 126:58–71, 2016.

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

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

A. Alesina, P. Giuliano, and N. Nunn. On the origin of gender roles: Women and the plough. *Quarterly Journal of Economics*, 128(2):469–530, 2013.

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

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.

A. Ben-Ner, B.P. Mccall, S. Massoud, and H. Wang. Identity and in-group/out-group differentiation in work and giving behaviors: Experimental evidence. *Journal of Economic Behavior and Organization*, 72(1):153–170, 2009.

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

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

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.

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

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.

J.P. Carvalho and B. Pradelski. Identity and underrepresenation: Interactions between race and gender. Working paper, Oxford University, 2019.

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.

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

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

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

E. Dekel, J.C. Ely, and O. Yilankaya. Evolution of preferences. *Review of Economic Studies*, 74:685–704, July 2007.

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

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

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

C.L. Exley and J.B. Kessler. The gender gap in self-promotion. *Quarterly Journal of Economics*, 137(3):1345–1381, 2022.

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

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

M. Frick, R. Iijima, and Y. Ishii. Dispersed behavior and perceptions in assortative societies. Discussion Paper 2650, Cowles Foundation, 2018.

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

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

S. Hodge, J.W. Burden, L.E. Robinson, and R. Bennett. Theorizing on the stereotyping of black male student-athletes: Issues and implications. *Journal for the Study of Sports and Athletes in Education*, 2(2):203–226, 2008.

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.

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

M. 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.

J.P. Leyens, M. Desert, J.C. Croizet, and C. Darcis. Stereotype threat: Are lower status and history of stigmatization preconditions of stereotype threat? *Personal. Soc. Psychol. Bull.*, 26(10):1189–1199, 2000.

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

D.M. Marx and P.A. Goff. Clearing the air: the effect of experimenter race on targets test performance and subjective experience. *Br. J. Soc. Psychol.*, 44(4):645–657, 2005.

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.

S. Oh. Does identity affect labor supply? Working paper, Paris School of Economics, 2019.

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

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

M. Peski and B. Szentes. Spontaneous discrimination. *American Economic Review*, 103 (6):2412–2436, 2013.

S. Peters. *A Path through the Jungle.* Mindfield Media, 2021.

T. Piketty. Social mobility and redistributive politics. *The Quarterly Journal of Economics*, 110(3):551–584, 1995.

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

V. Purdie-Vaughns, C.M. Steel, P.G. Davies, R. Ditlmann, and J.R. Crosby. Social identity contingencies: how diversity cues signal threat or safety for african americans in mainstream institutions. *Journal of Personal. Social Psychology*, 94(4):615–630, 2008.

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

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

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

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

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

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

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

J. Stone, C.I. Lynch, M. Sjomeling, and J.M. Darley. Stereotype threat effects on black and white athletic performance. *J. Personal. Soc. Psychol.*, 77(6):1213–1228, 1999.

Appendix 1: Proofs

PROPOSITION 1 (Individually Optimal Belief Formation): *The individually optimal strategies σ^* given an agent's type $\{\alpha, \theta\}$ are the following:*

- *The individually optimal strategy σ^* is 'Not Repress' for agents of type $\{\alpha, x\}$ such that $\alpha > \gamma$ and $\pi_x > p_x$ or $\alpha < \gamma$ and $\pi_x < p_x$*
- *The individually optimal strategy σ^* is 'Repress' for agents of type $\{\alpha, x\}$ such that $\alpha > \gamma$ and $\pi_x < p_x$ or $\alpha < \gamma$ and $\pi_x > p_x$*

Proof. Agents choose σ_i to maximize V_i over all possible realizations of $\hat{\alpha}_i$. Consider first agents that have an $\alpha > \gamma$. 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 $\sigma_i = R$. This means that we need $\Phi_{\alpha,x,NR,\Pi} \geq \Phi_{\alpha,x,R,\Pi}$. Since $\Phi_{\alpha,x,\sigma,\Pi} = P(\hat{\alpha} > \gamma^{\sigma}|\alpha)$, this is the case when $\gamma^{NR} < \gamma^{R}$. This is true if and only if $\pi_x \geq p_x$. Therefore, NR is only an optimal strategy for agents with $\alpha > \gamma$ when their observable characteristic $\theta_i = x$ is such that the social identity cue $\pi_x \geq p_x$. 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 $\sigma_i = R$. This means $\Phi_{\alpha,x,NR,\Pi} \leq \Phi_{\alpha,x,R,\Pi}$. This is the case if and only if $\gamma^{NR} > \gamma^{R}$, meaning that we need $\pi_x \leq p_x$. Therefore, agents with α should only choose strategy $\sigma_i = NR$, when their observable characteristic is such that $\pi_x \leq p_x$. Otherwise, they are better off choosing strategy $\sigma_i = R$. ■

PROPOSITION 2 (Existence Non-Neutral Regime): *Let $p_0 = p_1$, and let $\delta > 0$ be a small value with which we disturb a 'Neutral Regime'. A sufficient condition for the co-existence of a stable 'Non Neutral Regime' with a 'Neutral Regime' is as follows,*

$$\frac{\partial \gamma^{NR}}{\partial \delta} \frac{|\frac{\partial S}{\partial \gamma}|}{S} > 4 \tag{18}$$

where $\gamma^{NR} = \frac{\gamma}{\eta_x}$ and $S = \int \alpha G_\alpha(\gamma) f(\alpha) d\alpha$.

Proof. 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,

$$\tilde{\pi}_1(\pi_1, \sigma) = \frac{S_1}{S_1 + S_0} \tag{19}$$

where $S_1 = p_1 \int \alpha \Phi_{\alpha,1,\sigma,\Pi} f(\alpha) d\alpha$ and $S_0 = p_0 \int \alpha \Phi_{\alpha,0,\sigma,\Pi} f(\alpha) d\alpha$ denote the number of successful agents at the Competitive task with respectively $\theta = 1$ and $\theta = 0$. First of all, one can directly infer that $\pi_1 = p_1$ is an equilibrium, because in that case, the strategies NR and R are equivalent. Since α and θ are independently distributed, there will be no difference in choice behavior across different social groups, and $\tilde{\pi}_1(p_x, \sigma) = p_x$.

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 $\theta_i = 1$ and $\alpha > \gamma$ choose to undertake the *Competitive* task, and is equal to

$$\overline{S}_1 = p_1 \int_{\alpha<\gamma} \int_{\hat{\alpha}>\gamma} \alpha g_\alpha(\hat{\alpha}) f(\alpha) d\hat{\alpha} d\alpha + p_1 \int_{\alpha>\gamma} \alpha f(\alpha) d\alpha \tag{20}$$

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 $\theta_i = 0$ and $\alpha < \gamma$ choose to undertake the *Competitive* task, and is equal to

$$\underline{S}_0 = p_0 \int_{\alpha>\gamma} \int_{\hat{\alpha}>\gamma} \alpha g_\alpha(\hat{\alpha}) f(\alpha) d\hat{\alpha} d\alpha \tag{21}$$

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

$$\overline{\pi}_1(\sigma, \Pi) = \frac{\overline{S}_1}{\overline{S}_1 + \underline{S}_0} < 1 \tag{22}$$

It is now sufficient to show that, if $\frac{\partial \tilde{\pi}_1(\sigma,\Pi)}{\partial \pi_1}|_{\pi_1=p_1} > 1$, then there exists a stable 'Non-Neutral Regime'.

In a 'Neutral Regime', we have,

$$\frac{S_1}{S_0} = \frac{p_1}{p_0} \tag{23}$$

with $S_1 = p_1 \int \alpha G_\alpha(\gamma) f(\alpha) d\alpha$ and $S_0 = p_0 \int \alpha G_\alpha(\gamma) f(\alpha) d\alpha$. Let us now analyze the effect of a slight perturbation of a 'Neutral Regime', such that $\pi_1 = p_1 + \delta$, while $\pi_0 = p_0 - \delta$ and,

$$\frac{S_1}{S_0} > \frac{p_1}{p_0} \tag{24}$$

In this case, Proposition 1 shows that all agents with $\theta_i = 1$ and $\alpha > \gamma$ move from the choice rule in which $a = C$ when $\hat{\alpha} > \gamma$ to the rule $a = C$ when $\hat{\alpha} > \frac{\gamma}{\eta_1}$, while all agents with $\theta_i = 0$ and $\alpha < \gamma$ move from the choice rule in which $a = C$ when $\hat{\alpha} > \gamma$ to the rule $a = C$ when $\hat{\alpha} > \frac{\gamma}{\eta_0}$. Assume that the response function is continuous and such that agents process the difference $\pi_x - p_x$. Then,

$$\eta_1 = \eta(\pi_1 + \delta - p_1) = \eta(0) + \eta'(0)[\pi_1 + \delta - p_1] \tag{25}$$

Therefore,

$$\frac{\gamma}{\eta_1} = \frac{\gamma}{\eta(0) + \eta'(0)\delta} \tag{26}$$

$$\simeq \frac{\gamma}{\eta(0)}[1 - \frac{\eta'(0)}{\eta(0)}\delta] \tag{27}$$

Because $\eta(0) = 1$,

$$\frac{\gamma}{\eta_1} \simeq \gamma[1 - \eta'(0)\delta] \tag{28}$$

and we can write the change in the threshold γ^{NR} for agents moving from γ to $\frac{\gamma}{\eta_1}$ as,

$$\gamma - \frac{\gamma}{\eta_1} \simeq \gamma\eta'(0)\delta \tag{29}$$

This shows that this change is multiplicative in γ. Furthermore, let $\Delta_x = \gamma - \frac{\gamma}{\eta_x}$. Then, there is a symmetry, such that $\Delta_1 = -\Delta_0$. Let $S'_x = \frac{\partial S_x}{\partial \delta}$. Then,

$$S'_1 = S_1 + p_1 \int_{\alpha > \gamma} \alpha f(\alpha) d\alpha \int_{\gamma - \Delta_1 < \hat{\alpha} < \gamma} g_\alpha(\hat{\alpha}) d\hat{\alpha} \tag{30}$$

where $\int_{\gamma-\Delta_1<\hat{\alpha}<\gamma} g_\alpha(\hat{\alpha})d\hat{\alpha} \approx g_\alpha(\gamma)\Delta_1$. Similarly,

$$S_0' = S_0 - p_0 \int_{\alpha<\gamma} \alpha f(\alpha)d\alpha \int_{\gamma<\hat{\alpha}<\gamma-\Delta_0} g_\alpha(\hat{\alpha})d\hat{\alpha} \tag{31}$$

where $\int_{\gamma<\hat{\alpha}<\gamma-\Delta_0} g_\alpha(\hat{\alpha})d\hat{\alpha} \approx g_\alpha(\gamma)\Delta_0$. For Δ_θ arbitrarily small,

$$\frac{S_1'}{S_0'} = \frac{p_1}{p_0}\frac{S + \Delta_1 \int_{\alpha>\gamma} \alpha g_\alpha(\gamma) f(\alpha)d\alpha}{S - \Delta_0 \int_{\alpha<\gamma} \alpha g_\alpha(\gamma) f(\alpha)d\alpha} \tag{32}$$

$$= \frac{p_1}{p_0}\left[1 + \Delta\frac{\int_\alpha \alpha g_\alpha(\gamma) f(\alpha)d\alpha}{S}\right] \tag{33}$$

$$= \frac{p_1}{p_0}\left[1 + \Delta\frac{|\frac{\partial S}{\partial \gamma}|}{S}\right] \tag{34}$$

with $\Delta = \gamma\eta'(0)\delta$. A 'Neutral Regime' becomes unstable when,

$$\frac{p_1}{p_0}\left[1 + \Delta\frac{|\frac{\partial S}{\partial \gamma}|}{S}\right] > \frac{p_1+\delta}{p_0-\delta} \tag{35}$$

When $p_1 = p_0 = \frac{1}{2}$, this is the case when

$$\gamma\eta'(0)\delta\frac{|\frac{\partial S}{\partial \gamma}|}{S} > 4\delta \tag{36}$$

$$\gamma\eta'(0)\frac{|\frac{\partial S}{\partial \gamma}|}{S} > 4 \tag{37}$$

Finally, we note that $\gamma\eta'(0) = \frac{\partial(\gamma - \frac{\gamma}{\eta x})}{\partial \delta} = \frac{\partial \gamma^{NR}}{\partial \delta}$. This condition is also sufficient when $p_0 = p_1$, and agents process the proportion $\frac{\pi_x}{p_x}$. Note that,

$$\frac{\pi_1}{p_1} = \frac{p_1+\delta}{p_1} = 1 + \frac{\delta}{p_1} \tag{38}$$

and,

$$\frac{\pi_0}{p_0} = \frac{p_0+\delta}{p_0} = 1 - \frac{\delta}{p_0} \tag{39}$$

Then, when $p_1 = p_0$, $\Delta_1 = -\Delta_0$. Because $\Delta = \delta\eta'(0)\gamma$, locally we have,

$$\eta\left(\frac{\pi_x}{p_x}\right) \approx \eta(1 + 2(\pi_x - p_x)) \tag{40}$$

Therefore, in the case $p_0 = p_1$, the sufficient condition in (37) also applies when agents process the fraction $\frac{\pi_x}{p_x}$ instead of the difference $\pi_x - p_x$. ■

COROLLARY 1: *Let $x' \in \{0,1\}$ be the complement of x and assume WLOG that $\pi_x > \pi_{x'}$. The optimal use of social identity has both a population effect, such that $\Phi_{\alpha,x,\sigma_i,\Pi} > \Phi_{\alpha,x',\sigma_i,\Pi}$ and a selection effect, such that $E(\alpha|a = C, x) < E(\alpha|a = C, x')$. The strength of both effects is such that the order $\pi_x > \pi_{x'}$ will always be preserved.*

Proof. Assume WLOG that $\pi_1 > \pi_0$. Then, $\frac{\gamma}{\eta_1} < \gamma$, while $\frac{\gamma}{\eta_0} > \gamma$. Therefore, all agents with $\alpha > \gamma$ and $\theta_i = 1$ will choose $\gamma_i = \frac{\gamma}{\eta_1}$, 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 $\frac{\gamma}{\eta_1} < \gamma$, we know that agents with $\theta_i = 0$ 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, we can show that,

$$\frac{\partial \tilde{\pi}_1(\pi_1, \sigma)}{\partial \pi_1} = \frac{\pi_1 \left[\frac{\partial S_1}{\partial \pi_1} - \frac{\partial S_0}{\partial \pi_1}\right]}{S} \tag{41}$$

where $S_1 = p_1 \int \alpha \Phi_{\alpha,1,\sigma,\Pi} f(\alpha) d\alpha$ and $S_0 = p_0 \int \alpha \Phi_{\alpha,0,\sigma,\Pi} f(\alpha) d\alpha$ denote the number of successful agents at the Competitive task with respectively $\theta_i = 1$ and $\theta_i = 0$. Furthermore, $\frac{\partial S_1}{\partial \pi_1}$ for $\pi_1 \in [p_1, 1]$ is given by,

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

Similarly,

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

Therefore, $\frac{\partial S_1}{\partial \pi_1} > 0$, 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: *Take two response functions $\hat{\eta}$ and η, such that $\hat{\eta}(\pi_x, p_x) > \eta(\pi_x, p_x)$ for all $\pi_x > p_x$. 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 response function η. Then, $\pi^*_{\hat{\eta},1} > \pi^*_{\eta,1}$.*

Proof. Assume $\eta(\pi_x, p_x)$ is a response 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. Then, it follows from Proposition 2 that a 'Non-Neutral Regime' also exists for any response function $\hat{\eta}(\pi_x, p_x)$, such that $\hat{\eta}(\pi_x, p_x) > \eta(\pi_x, p_x)$. Let $\tilde{\pi}_{\eta,1}(\pi_1, \sigma)$ be the induced value of π_1 for a response function η. Then, if $\hat{\eta}(\pi_1, p_1) > \eta(\pi_1, p_1)$ for all $\pi_1 > p_1$,

$$\tilde{\pi}_{\hat{\eta},1}(\pi_1, \sigma) > \tilde{\pi}_{\eta,1}(\pi_1, \sigma) \quad \forall \pi_1 > p_1 \tag{42}$$

Consequently, let $\pi^*_{\eta,1}$ be the equilibrium value of π_1 that arises in a 'Non-Neutral Regime' for a response function η. Then,

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

which implies that,

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

and,

$$\pi^{(3)} \equiv \tilde{\pi}_{\hat{\eta},1}(\pi^{(2)}, \sigma) > \tilde{\pi}_{\hat{\eta},1}(\pi^{(1)}, \sigma) \equiv \pi^{(2)} \tag{45}$$

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 response function $\hat{\eta}(\pi_x, p_x)$ such that $\hat{\eta}(\pi_1, p_1) > \eta(\pi_1, p_1)$ for all $\pi_1 > p_1$, in equilibrium

$$\pi^*_{\hat{\eta},1} > \pi^*_{\eta,1} \tag{46}$$

■

COROLLARY 3: *Assume WLOG that agents hold a belief $\hat{p}_x > p_x$. Then there only exists a 'Non-Neutral Regime' in which $\pi_x < p_x$. A 'Neutral Regime' no longer exists.*

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

$$\eta(\pi_0, p_0) = \begin{cases} > 1 & \text{if } \pi_0 > \hat{p}_0 \\ 1 & \text{if } \pi_0 = \hat{p}_0 \\ < 1 & \text{if } \pi_0 < \hat{p}_0 \end{cases} \tag{47}$$

and consequently, when $\pi_0 = p_0$, $\eta_0 < 1$. 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 $\pi_0 < p_0$. ■