Stereotypes, Underconfidence and Decision-Making with an Application to Gender and Math*

Elyès Jouini[†] Paul Karehnke [‡] Clotilde Napp[§]

February 7, 2018

Abstract

We study the effects of the presence of a negative stereotype on the formation of self-confidence and on decision-making in achievement-related situations. We take into account not only consumption utility but also psychological utility (ex-ante ego utility and ex-post disappointment/elation). We show that any stereotype of lower ability (in the form of biased interpretation of success and failure in terms of ability) leads to gaps in confidence, in participation in risky/ambitious options and in performance. Furthermore, we show how the stereotype survives and even gets reinforced. Considering gender and mathematics, we are able to explain the lower self-confidence of girls in mathematics, their underrepresentation in STEM fields, as well as their choices of less ambitious options and lower performance.

Keywords: stereotype, gender gap, self-confidence, subjective ability.

JEL: D03, D81, D84, J16, I24

*We thank two anonymous referees, Renée Adams, Sebastian Ebert, Gigi Foster and Sudipta Sarangi (the Editor) as well as conference participants of M-BEES 2015, the Women in Leadership Conference at UNSW (2015) and the 15th Journées LAGV for very helpful comments and suggestions. The financial support of the GIP-ANR (2010 BLANC 1803 01 RISK project) and of the Risk Foundation (Groupama Chair) is gratefully acknowledged by Elyès Jouini and Clotilde Napp. Paul Karehnke gratefully acknowledges financial support from the Women in Leadership research network of the UNSW Business School.

[†]PSL, Univ. Paris-Dauphine, CEREMADE, e-mail: jouini@ceremade.dauphine.fr.

[‡]School of Banking and Finance, The University of New South Wales, e-mail: p.karehnke@unsw.edu.au.

[§]CNRS and PSL, Univ. Paris-Dauphine, e-mail: clotilde.napp@dauphine.fr.

1 Introduction

The underrepresentation of girls in STEM fields is substantial and poses a problem of both equity and efficacy, especially as jobs in these fields are among the most highly paid.[1] This gender gap in choices of STEM participation persists controlling for ability, which excludes fully ability-based explanations. As noted in Niederle and Vesterlund [2010, p. 141], "among equally gifted students, males are many times more likely to select college majors that are considered to be high in math content."[2] The main explanation provided in the literature for this gender gap in choices is a gender gap in self-confidence: with equivalent results, girls are less optimistic about their abilities in mathematics [Correll, 2001]. This first raises the question about the origin of this underconfidence. Second, because girls need to be more able to reach the same level of confidence as boys, if underconfidence were the only reason for lower self-selection, only the most able would self-select. As a consequence, the gender gap in choices would be lower for high ability levels, and self-selected girls would perform better than self-selected boys, which is not observed in the data. Finally, there is a gender gap in performance in math, at the advantage of boys, especially at high ability levels.[3]

This paper aims at providing a decision model that permits to explain these three gaps (in self-confidence, in choices, and in performance) as well as their persistence without assuming differences in objective ability. In particular, such a model needs to propose a mechanism of self-confidence formation, leading girls to be less self-confident (controlling for objective ability), as well as a mechanism of decision-making leading girls to self-select less (for the same level of objective ability). Furthermore, these mechanisms should not lead to the self-selection of only the most able girls.

We propose a model of (optimal) belief formation and choices that draws on models with ex-ante savoring and ex-post disappointment [in particular Gollier and Muermann, 2010][4] and

[1] E.g., Kirkeboen et al. [2016] estimate that individuals choosing science instead of humanities as college fields almost triple their early career income. In addition, Joensen and Nielsen [2016] identify a causal effect of mathematics on gender differences in career outcomes.

[2] See also Weinberger [2005] who shows that white women enter math-related fields at no more than half the rate of men with the same mathematics test scores. In Buser et al. [2014], students exhibit a significant gender gap in math intensity of their chosen profiles, controlling for objective academic performance.

[3] In the Programme for International Student Assessment (PISA) there are on average among OECD countries 40% of girls among top performers in math [Guiso et al., 2008]. At the SAT, boys outperform girls in math on average and girls represent 30% of top performers [Ellison and Swanson, 2010].

[4] See also Loewenstein and Linville [1986], Karlsson et al. [2004], Jouini et al. [2014], and Macera [2014]. Although less common in the economics literature, this model explains observed violations of expected utility such as the preference reversal in the Allais paradox [Gollier and Muermann, 2010] and simultaneous demand for lotteries and

extends them to take into account stereotype issues. Individuals derive not only consumption utility but also psychological utility in the form of ego utility ex-ante and disappointment/elation ex-post. Ego utility relies on the self-confidence level and ex-post disappointment and elation are measured relative to ex-ante self-confidence. Optimal self-confidence realizes the best trade-off between today's ego utility and tomorrow's elation or disappointment; it represents the optimal self-insurance level against future psychological risk of disappointment/elation.

We extend this optimal belief formation mechanism with a stereotype component. We assume that boys and girls have the same distribution of ability. However, building on status characteristics theory and expectation states theory [Berger et al., 1972, Foschi, 2000, Ridgeway, 2001], we assume that the presence of a stereotype, in the form of a biased/stereotyped attribution of success and failure in terms of ability, exposes boys and girls to different psychological risk. These differences in psychological risk can be interpreted as differences in self-esteem risk and generate different protection strategies in terms of self-confidence and choices. The underconfidence of girls as well as their less ambitious/challenging, less risk tolerant or competitive choices are their best response to their higher psychological risk. By adopting modest confidence levels and by choosing less ambitious options, girls reduce their future psychological risk of disappointment if reality does not keep up with expectations. Besides, we show that differences in psychological risk combined with our self-confidence formation mechanism generate differences in the nature of the relation between ability and self-selection in difficult tasks. The relationship is more increasing for boys than for girls, in the sense that more able boys self-select more in difficult options whereas highly able girls might self-select less than less able girls.

This gender selection effect in choosing challenging options leads to higher ability conditional on participation, i.e., better performance of boys, especially for high ability levels. This self-selection bias is an important feature of our approach; even if the same (ex-ante) ability is assumed, the mere presence of the stereotype leads to girls underperformance. This feature permits the survival of the stereotype and provides a possible channel through which status hierarchies can survive and even get reinforced.

In our model, the presence of a negative stereotype about girls and math can explain the three gender gaps (in confidence, participation, performance) as well as their main features. The same

insurance [Jouini et al., 2014].

approach can be applied to gender issues in general because there is a stereotype of lower ability of women in many fields [Foschi, 1996, Fiske et al., 2002]. More generally, the same approach can be applied to any negatively stereotyped group in achievement-related situations, as long as there is evidence of a biased interpretation of success and failure in terms of ability, leading to underconfidence, avoidance of the stereotyped field, less ambitious choices, underrepresentation in most selective tracks, possible underperformance, and to the survival of the stereotype.

Our results strongly rely on the self-insurance strategy linked to ex-ante savoring and ex-post elation/disappointment. In particular, a model with the negative stereotype being represented by higher psychological risk in the form of higher costs and lower benefits (independent of expectations) or with psychological risk dependent on objective beliefs a la Bell [1985] or Kőszegi and Rabin [2006] would only lead to a gap in choices of risky/ambitious options. It would provide no result on self-confidence and would lead to the overperformance of the stereotyped group (because only the most able would self-select).

Our approach shares similarities with Mechtenberg [2009], one of the few theoretical approaches to the issue of gender and mathematics, which also leads to boys outperforming girls in math and science and to the predominance of male students in math majors at university. However, while Mechtenberg [2009] relies on biased grading and on a cheap talk model of teachers and students, we rely on a stereotype model. Our analysis provides socio-cultural foundations for the gender gaps in mathematics and for their survival. It is in line with Correll [2001], who measures the extent to which cultural beliefs about gender and mathematics bias the formation of self-assessments of competence and contribute to the gender gap in careers in science, mathematics, and engineering.

Our approach is different from the recent work of Bordalo et al. [2016] on stereotypes. In their approach, stereotypes emerge and some groups are considered and consider themselves less able, hence participate less in difficult options, because of an exaggeration of differences in objective ability distributions. In our approach, we assume no ex-ante difference in ability, and do not analyze how the stereotype emerges but rather focus on its implications in terms of self-confidence and decision-making as well as how it survives.

The paper is organized as follows. Section 2 presents the general model of stereotypes and discusses its assumptions. Section 3 contains the results (self-confidence, choices, and performance) as well as numerical calculations. Section 4 discusses applications to gender and math as well as other stereotypes, and concludes. All proofs are in Appendix A.

2 Model

We rely on models with ex-ante savoring and ex-post disappointment like Karlsson et al. [2004], Gollier and Muermann [2010], and Jouini et al. [2014] but add in a stereotype component. We consider an individual who is confronted with risky situations involving her ability. There are two dates, denoted by date 0 and date 1. A risky situation is represented by a random variable $\widetilde{x}$ with two possible outcomes at date 1: $\widetilde{x} = x_l$, representing failure, and $\widetilde{x} = x_h$, representing success, with $0 \leq x_l \leq x_h \leq 1$. We can think of $\widetilde{x}$ as representing graduation risk, for instance, or any pass or fail exam. We let p denote the individual's objective probability of success; the individual's objective expectation of performance is then given by $E[\widetilde{x}] = px_h + (1-p)\,x_l \in [x_l, x_h]$. We assume that the individual's perception of the probability of success might differ from the objective one, and we let $y \in [x_l, x_h]$ denote the individual's subjective expectation of performance. We shall also refer to y as the individual self-confidence level.

At date 0, the individual has anticipatory ego utility [Kőszegi, 2006, Weinberg, 2009], in the form of an increasing function $v(y)$ of her subjective expectation of performance. Indeed, a high subjective expectation of performance increases the individual's satisfaction at date 0 by increasing her feeling of personal capacity. At date 1, the individual's utility depends upon the realized outcome $x \in \{x_l, x_h\}$, upon the expectation of performance y and upon the intensity of the negative stereotype $\lambda \in [0,1]$ and is given by $u(x) + \varphi(x, y, \lambda)$. The first component $u(x)$ is standard 'consumption' or outcome utility. The second component $\varphi(x, y, \lambda)$ represents ex-post psychological utility, stemming from elation or disappointment. We let $W(y)$ denote the individual's intertemporal well-being

$$W(y) = v(y) + E[u(\widetilde{x})] + E[\varphi(\widetilde{x}, y, \lambda)]\,. \tag{1}$$

The difference between (1) and previous savoring and disappointment models [Karlsson et al., 2004, Gollier and Muermann, 2010, Jouini et al., 2014] stems from the dependence in λ, representing the impact of stereotypes. We make the following assumptions on the components of well-being.

Assumption (A1): *u is increasing and strictly concave with $u_{yy} < 0$, $u(0) = 0$ and $v = ku$ for*

$k \in \mathbb{R}_+^*$.

Assumption (A2): $\varphi(x_l, y, \lambda) \leq 0, \varphi(x_h, y, \lambda) \geq 0$.

Assumption (A3): $\varphi_y \leq 0, \varphi_{yy} \leq 0$.

Assumption (A4): $\varphi_y(x_h, y, \lambda) \geq \varphi_y(x_l, y, \lambda)$.

Assumption (A5): $\varphi_\lambda \leq 0$.

Assumption (A6): $\varphi_{y\lambda}(x_l, y, \lambda) \leq 0, \varphi_{y\lambda}(x_h, y, \lambda) \geq 0$.

The assumptions on u in Assumption (A1) are standard. The assumption on $u(0)$ is a simplifying assumption. So is the assumption on v, that permits to analyze the impact of the weight on ego feelings.[5] Assumption (A2) amounts to assuming that success has a positive psychological impact and that failure has a negative one. Assumption (A3) is an elation/disappointment condition: the ex-ante expectation y plays the role of a reference level and the higher the ex-ante expectations, the lower the ex-post psychological benefit.[6] This ex-post psychological utility stems from the comparison between expectations and the realized outcome (success or failure). It can be interpreted as ex-post self-esteem; according to James [1890], Diener et al. [1991], and Mellers and McGraw [2001], self-esteem is the relation between one's real self and one's ideal self, hence increases with one's successes and decreases with one's expectations. The additional condition $\varphi_{yy} \leq 0$ is a regularity condition ensuring the concavity of the well-being function $W(y)$ as in models of savoring and disappointment [Gollier and Muermann, 2010, Jouini et al., 2014]. Moreover, we make the Assumption (A4) that disappointment effects are more important than elation effects. There is strong empirical support for this assumption [see, e.g., Mellers et al., 1997, Mellers and McGraw, 2001].

The last two assumptions are about the impact of stereotypes. We assume that stereotypes impact the way success and failure are interpreted in terms of ability. According to status characteristics theory and expectation states theory [Berger et al., 1972, Foschi, 2000, Ridgeway, 2001], an individual who is considered less able (i.e., a negatively stereotyped individual or a low-status individual) will have failures attributed to lack of ability and successes attributed externally—to luck, for example.[7] The reason is that for an individual who is expected to be less able, failure is

[5]Notice that the assumption on v ensures that in the absence of risk our intertemporal well-being simplifies to the intertemporal well-being in anticipation models without risk a la Loewenstein [1987].

[6]See Bell [1985], Loomes and Sugden [1986], Gul [1991] for models of disappointment, and Kőszegi and Rabin [2006] for reference-dependent models.

[7]Foschi [2000] for instance argues that the standards used to determine if a given performance is indicative of

consistent with expectations of ability but success is not. The opposite holds for an individual who is considered as more able; they will have successes attributed to ability and failures attributed externally—to lack of luck. In line with these arguments, Assumption (A5) assumes that when the negative stereotype is stronger, there is less elation from success (because it is more externally attributed) and there is more utility loss from failure (because it is more internally attributed). In addition, Assumption (A6) assumes that when the negative stereotype is higher, there is less marginal elation from success and more marginal disappointment from failure. Let us illustrate these assumptions on a natural extension of Bell [1985]'s model[8] of disappointment and elation to our setting with stereotypes, with $\varphi(x_h, y, \lambda) = (1-\lambda) K (x_h - y)$, $\varphi(x_l, y, \lambda) = -\lambda\eta(y - x_l)$ for $(K, \eta) \in \mathbb{R}_+^2$. We have $\varphi_\lambda \leq 0$, $\varphi_{y\lambda}(x_h, y, \lambda) = K \geq 0$, $\varphi_{y\lambda}(x_l, y, \lambda) = -\eta \leq 0$ and Assumptions (A5) and (A6) are satisfied.

We know by Assumption (A5) that the psychological benefit from success decreases with the negative stereotype and that the loss from failure increases with the stereotype. We normalize the maximal and minimal stereotype levels λ =1 and $\lambda = 0$ such that there is no benefit from success for $\lambda = 1$ and no loss from failure for $\lambda = 0$, i.e., $\varphi(x_l, y, 0) = 0$ and $\varphi(x_h, y, 1) = 0$.

In our model, the individual faces psychological risk $\varphi(\widetilde{x}, y, \lambda)$ at date 1, in addition to the standard consumption risk $u(\widetilde{x})$. Under our assumptions, controlling for y, this risk is higher (in the sense of first-order stochastic dominance) for individuals with higher levels of λ, i.e., the negative stereotype increases date 1 psychological risk.[9] Notice, however, that date 1 psychological risk increases with y, hence negatively stereotyped individuals may still experience less disappointment from failure and substantial elation from success due to lower individual expectations of performance (or self-confidence levels).

As in Brunnermeier and Parker [2005], we assume that individuals optimally choose their subjective expectations [see also Akerlof and Dickens, 1982, Caplin and Leahy, 2001, Kopczuk and Slemrod, 2005]. More precisely, we assume that the individual adopts the subjective expectation

ability are a function of the status of the individual.

[8]Bell [1985] considers two possible outcomes x_l and x_h with $x_l < x_h$. Letting y denote the expectation, disappointment felt in x_l is given by $\eta(y - x_l)$ and elation felt in x_h is given by $K(x_h - y)$ for non-negative constants η and K.

[9]We recall the definition of first-order stochastic dominance. A random variable X dominates a random variable Y in the sense of FSD if and only if the cumulative distribution functions satisfy $F_X \leq F_Y$.

of performance y^* in $[x_l, x_h]$ that maximizes her intertemporal well-being $W(y)$, i.e., solves

$$y^* \equiv \arg \max_{y \in [x_l, x_h]} W(y) = \arg \max_{y \in [x_l, x_h]} v(y) + E[\varphi(\widetilde{x}, y, \lambda)]. \tag{2}$$

The individual facing the risky situation $\widetilde{x}$ is then endowed with the well-being level $W(y^*)$. The individual's optimal expectation of performance or optimal self-confidence realizes the best trade-off between today's ego utility and tomorrow's disappointment/elation. It characterizes the level of ex-ante ego utility that the individual is ready to sacrifice to reduce future psychological risk—i.e., the optimal self-insurance level against future self-esteem risk. Note that in our model, as in Akerlof and Dickens [1982], Brunnermeier and Parker [2005], and Gollier and Muermann [2010], individuals use *objective* probabilities to evaluate future utility while experiencing ego utility related to *subjective* expectations; indeed, optimal levels of confidence are those that maximize the individual's satisfaction on average across realizations of uncertainty and uncertainty unfolds according to objective probability. The individual faces some cognitive dissonance [Festinger, 1957]. This is in line with experiments in psychology [see Gollier and Muermann, 2010, for a discussion]. As an illustration, consider the case of a student who takes an important pass or fail exam. The student's intertemporal well-being consists both of her satisfaction before uncertainty resolves, i.e., her ego utility, and of her satisfaction after uncertainty resolves, once she knows her results. Either she enjoys a high level of self-confidence and benefits from a good ego feeling ex-ante, but this comes at the risk of experiencing disappointment ex-post, if reality is below expectations. Or she adopts a low level of self-confidence, which is associated with less ego utility ex-ante, but comes at the benefit of being less exposed to future loss of self-esteem.

Note that if there is no uncertainty involved, i.e., if $\widetilde{x} = x_l = x_h = A$, then according to (A2), we have $\varphi = 0$, i.e., there is no psychological risk, and according to (2), we have $y^* = A$, i.e., there is no ex-ante manipulation of self-confidence. Well-being is then given by $W(y^*) = (k+1)\,u(A)$; we denote it by W_A.

3 Results

3.1 Self-confidence

Due to Assumptions (A1) and (A3), the well-being function is concave, and the first order conditions characterize the optimal self-confidence level y^*, which is given by:

- $y^* = x_l$ if $p \leq \frac{-\varphi_y(x_l,x_l,\lambda)-v_y(x_l)}{[\varphi_y(x_h,x_l,\lambda)-\varphi_y(x_l,x_l,\lambda)]}$,
- $y^* = x_h$ if $p \geq \frac{-\varphi_y(x_l,x_h,\lambda)-v_y(x_h)}{[\varphi_y(x_h,x_h,\lambda)-\varphi_y(x_l,x_h,\lambda)]}$,
- $v_y(y^*) + p\varphi_y(x_h, y^*, \lambda) + (1-p)\varphi_y(x_l, y^*, \lambda) = 0$ otherwise.

Proposition 1 *1a. Optimal self-confidence weakly increases with ability, i.e.,* $\frac{\partial y^*}{\partial p} \geq 0$.

1b. If the weight on ego utility (or, more precisely, the marginal ego utility $v_y(x_h)$) is large enough, then optimal self-confidence weakly decreases with the negative stereotype, i.e., $\frac{\partial y^*}{\partial \lambda} \leq 0$.

2. Individual well-being weakly increases with objective ability, i.e., $\frac{\partial W(y^*)}{\partial p} \geq 0$, *and weakly decreases with the negative stereotype, i.e.,* $\frac{\partial W(y^*)}{\partial \lambda} \leq 0$.

Although not immediate in our model, we get the desirable feature that self-confidence increases with ability as described in Proposition 1.1a. Because disappointment effects are more important than elation effects by Assumption (A4), as p increases, expected psychological costs decrease and the decision maker chooses higher self-confidence levels. Concerning the impact of the stereotype in Proposition 1.1b, we need an additional condition on the marginal ego utility to ensure that a higher level of the negative stereotype (that is associated with higher date 1 psychological risk) leads to a lower self-confidence level for all levels of ability. Indeed, as shown in the proof of the proposition, self-confidence always decreases with the intensity of the stereotype for levels of ability below a given threshold, but for higher levels of ability, the effect of success is dominant, and an increase in λ is associated with lower elation and hence may lead to a higher y^*. The condition on v_y rules out this mechanism. Taken together, Proposition 1.1a and 1.1b imply that the argument (mentioned in the introduction) about how negatively stereotyped individuals need to be more able to be as confident holds.

In Proposition 1.2 an increase in the ability level, or an increase in the level of the stereotype, have the expected impact on individual well-being. First, a higher level of ability increases well-being because it increases consumption utility due to Assumption (A1); it increases ex-post psychological utility due to Assumption (A2); and it permits higher levels of ego utility (without fearing exposure to loss of self-esteem). Second, higher stereotype levels lead to lower well-being because they are associated with higher costs of failure and lower benefits from success due to Assumption (A5). This result implies that individuals will be underrepresented in the fields where they are negatively stereotyped. Indeed, for the same level of ability, and the same consumption utility, an individual will choose the fields that maximize her well-being. According to Proposition 1.2, these fields are those in which she is less negatively stereotyped. Let us now consider choices between more or less challenging options within a given field.

3.2 Choice

We analyze the decision to participate in a challenging task—i.e., a risky situation involving one's ability (like the choice of a difficult track). We assume that the individual has the choice between the risky option $\widetilde{x}$ and a non-risky option. The non-risky option may depend upon the individual's ability and we denote it by $A(p) \in (x_l, x_h)$. To choose between these options, the decision maker first chooses her optimal expectation y^* for the risky option; for the non-risky option, the optimal expectation is given by the sure outcome. Well-being for the risky option is given by $W(y^*)$ and well-being for the non-risky option is given by $W_{A(p)} = (k+1)\,u(A(p))$. The individual chooses the option that maximizes her intertemporal well-being, i.e., participates in the risky option if and only if $W(y^*) \geq W_{A(p)}$, or equivalently if and only if

$$V(p,\lambda) \equiv [v(y^*) - v(A(p))] + [pu(x_h) + (1-p)\,u(x_l) - u(A(p))] + E[\varphi(\widetilde{x}, y^*, \lambda)] \geq 0, \qquad (3)$$

where $V(p,\lambda) \equiv W(y^*) - W_{A(p)}$ denotes the value of participation. Letting as usual the certainty equivalent $CE_{\widetilde{x},u}$ denote the constant such that $u\left(CE_{\widetilde{x},u}\right) = E[u(\widetilde{x})]$, the second component $pu(x_h) + (1-p)\,u(x_l) - u(A(p))$ can equivalently be written as $u\left(CE_{\widetilde{x},u}\right) - u(A(p))$. In the standard setting (corresponding to the case $v = \varphi = 0$ in our setting), the individual compares expected 'consumption utility levels' in both situations and chooses the risky option if and only if $CE_{\widetilde{x},u} \geq A(p)$. In our setting, two other components must be taken into account for decision-making. The first component, $v(y^*) - v(A(p))$, corresponds to the gain or loss in ego utility at date 0. It is in favor of the non-risky option if the subjective expectation y^* is below the non-risky outcome, i.e., if $y^* < A(p)$. The other component, $E[\varphi(\widetilde{x}, y^*, \lambda)]$, corresponds to the expected psychological gain or loss at date 1. It is in favor of the non-risky option if the average psychological impact of the risky option is negative.

Because the non risky option depends upon the ability level, the choice itself depends upon how $A(p)$ varies with p. We make the following assumption.

Assumption (B): *$u(A(p))$ is concave in p and $A(1) = x_h$.*

If A is increasing in p, the concavity assumption amounts to assuming that $\frac{A''(p)}{A'(p)} \leq -\frac{u''[A(p)]}{u'[A(p)]}$. Natural examples are given by $A(p) = px_h + (1-p)x_l$ (the objective expectation) or $A(p) = CE_{\widetilde{x},u}$. Assumption (B) in particular is always satisfied if A is concave in p.

As an illustration for our choice setting, consider a high school graduate in France who hesitates between preparatory class to elite school (CPGE) and university. CPGE is (seen as) more prestigious, more difficult and risky. It leads to a difficult competitive exam, threatening one's self-esteem. University on the contrary is seen as less prestigious, graduation is not competitive and grades essentially reflect the student's ability. The random variable $\widetilde{x}$ represents success or failure at entering an elite school and p represents the individual's objective probability of success. Well-being for the option CPGE is given by $W(y^*)$ and well-being for the option university is given by $W_{A(p)}$. The choice involves the consumption values of both options (in terms of future career opportunities for instance) but also immediate psychological rewards (CPGE might be more ego satisfying because it is seen as being more prestigious) and future self-esteem consequences (failure in CPGE can be disappointing and painful but success can be self-esteem rewarding).

Proposition 2 *1. Participation in the risky option weakly decreases with the negative stereotype: the participation set $S(\lambda) = \{p : V(p, \lambda) \geq 0\}$ weakly decreases with λ, in the sense of the set inclusion, i.e., $S(\lambda_1) \subseteq S(\lambda_2)$ for $\lambda_1 \geq \lambda_2$.*

2. There exists a threshold $\widehat{p}(\lambda)$ in $[0, 1]$ such that the value of participation in the risky option $V(p, \lambda)$ weakly decreases with ability p on $[0, \widehat{p}(\lambda))$ and weakly increases on $[\widehat{p}(\lambda), 1]$. The participation set $S(\lambda) = \{p : V(p, \lambda) \geq 0\}$ is of the form $[0, p_0(\lambda)] \cup [p_1(\lambda), 1]$ for some $p_0(\lambda)$

and $p_1(\lambda)$ *in* $[0,1]$.

As shown in Proposition 2.1 and due to Assumption (A5), a higher level of negative stereotyping leads to a lower participation. More negatively stereotyped individuals face a higher self-esteem risk and not only choose more modest confidence levels (Proposition 1) but also make more timid choices. Hence, for a given level of ability, participation in the risky option decreases with the negative stereotype. As an immediate consequence of Propositions 1 and 2.1, controlling for ability, and if the weight on ego feelings is large enough, then participation in the risky option weakly increases with the level of self-confidence y^*. But as shown in Proposition 2.2, participation in the challenging option is not always increasing with the level of ability p. For $\widehat{p}(\lambda)$ in $(0,1)$, an increase in p raises the incentive to participate only for high enough levels of ability. The reason is that for high ability levels, the increase in self-esteem rewards due to the higher probability of success are greater than the increase in ego rewards in the non-risky option, which is not necessarily the case for low ability levels, which are associated with lower self-esteem rewards. As a consequence, the individuals who self-select in challenging options are not necessarily the most able. The participation set is not restricted to the most able and may also include the least able; indeed, the least able may also participate because they have "less to lose."

The following result shows that strong negative stereotypes not only reduce participation as shown in Proposition 2, but also modify the nature of the relationship between ability and participation, the relationship being "less increasing" in the following sense.

Corollary 3 *1. The derivative of the value of participation with respect to objective ability* $V_p(p,\lambda)|_{V=0}$ *weakly decreases with* λ *if* $v_y(x_h)$ *is large enough.*

2a. If the negative stereotype is maximal, then participation weakly decreases with ability.

2b. If the negative stereotype is low enough and the self-esteem reward from success $\varphi(x_h, x_l, \lambda)$ *is above a given level, then participation weakly increases with ability.*

As seen in Corollary 3.1, a greater negative stereotype reduces the incentive to participate when ability increases. The reason is that a stronger stereotype reduces the psychological rewards associated with a higher probability of success. In the extreme, these rewards are eliminated and so is the incentive to participate. Note that the derivative $V_p|_{V=0}$ characterizes the impact of an increase in ability on the pivotal individual, that is, an individual who is indifferent between participating or not.

As shown in Corollary 3.2, a maximal negative stereotype biases self-selection in such a way that participation decreases with ability; more able individuals then self-select less. A high enough gap in the level of the stereotypes leads to a participation gap that is higher for the highly able than for the less able. This self-selection bias is important for performance issues.

3.3 Performance

We have seen that negatively stereotyped individuals are less confident (under an additional condition on the marginal ego utility), that they need to be more able to be as confident and that they participate in challenging options less. The argument that only the most able stereotyped individuals participate is not valid, as the following proposition shows.

Proposition 4 *Consider two groups G_1 and G_2 with stereotype parameters $\lambda_1 > \lambda_2$ and with the same distribution of ability.*

1. *If λ_1 is high enough, then the average ability of the individuals of Group G_1 participating in the risky option is lower than that of group G_2.*
2. *In the setting of Bell (1985)'s model of disappointment/elation, with quadratic utility functions and with $A(p) = CE_{\widetilde{x},u}$, then for all stereotype parameters $\lambda_1 > \lambda_2$, Group G_1 has a lower ability conditional on participation than Group G_2.*

Due to the aforementioned self-selection bias, a high enough gap in the stereotype levels leads in the general case to a gap in performance (i.e., on ability conditional on participation). We emphasize that the only difference between groups G_1 and G_2 is a difference in the level of the stereotype. Moreover, this result remains true if we restrict our attention to levels of ability above a given threshold, which means that our results not only help explain the overall lower performance of negatively stereotyped groups but can account for their underrepresentation at the right tail of the performance distribution, without assuming any difference in ability.

Note that in addition to the lower ability conditional on participation due to the self-selection bias, our model also provides indirect effects that possibly contribute to the lower performance of negatively stereotyped individuals. First, as seen in Proposition 1, they are less confident in their abilities, which may lead to performance-hindering anxiety, lower effort and lower persistence. Second, as seen in Proposition 1.2 and Proposition 2, stereotyped individuals avoid the negatively stereotyped fields and self-select less in difficult courses or options, which may lead to lower competence and hence to lower performance.

This result on the performance gap helps explain how stereotypes survive and can even get reinforced. Indeed, consider a (natural) dynamic on the intensity of the negative stereotype such that the stereotype level of a given group evolves positively following relative success of the group, and negatively following relative failure. As a consequence of Proposition 4, a high enough initial gap in the stereotype levels leads to an increase of the gap—the more negatively stereotyped group becoming more negatively stereotyped and the more positively stereotyped group becoming more positively stereotyped.

3.4 Specifications and numerical results

To illustrate our propositions, we conduct numerical calculations. We set $\{x_l, x_h\} = \{0, 1\}$ and consider:

Example 1 [Bell, 1985]: $\varphi(1, y, \lambda) = (1-\lambda) K (1-y)$, $\varphi(0, y, \lambda) = -\lambda \eta y$ with $(K, \eta) \in \mathbb{R}_+^2$.

Example 2: $\varphi(1, y, \lambda) = (1-\lambda) K$, $\varphi(0, y, \lambda) = -\lambda \eta y$ with $(K, \eta) \in \mathbb{R}_+^2$.

It is easy to verify that Example 2 satisfies all Assumptions (A2)-(A6). Example 1 is Bell's model introduced above. It satisfies Assumptions (A2)-(A3) and (A5)-(A6) and further satisfies Assumption (A4) when $\lambda \geq \frac{K}{\eta+K}$. With quadratic utility functions $u(x) = x - \frac{1}{2}\alpha x^2$ and $v = ku$, these examples satisfy Assumption (A1). In addition, the condition on $v_y(1)$ in Proposition 1 is always satisfied in the setting of Example 2 and it is given by $ku_y(1) \geq \frac{\eta K}{(\eta+K)}$ in the setting of Example 1. Let Examples 1Q and 2Q denote the setting with psychological utility functions of Examples 1 and 2 with quadratic (outcome) utility functions and with $A(p) = CE_{\widetilde{x},u}$.

In Appendix A, we provide explicit characterizations of optimal self-confidence levels and of choices (participation in the risky option) in the setting of Examples 1Q and 2Q. These explicit results permit numerical simulations. We consider two groups, "boys" and "girls," with the same uniform distribution of objective ability on $[0, 1]$. In Example 1Q, we use the specification (α, k, η, K) = (0.8, 1.25, 1, 0.3),[10] and negative stereotype levels of $\lambda = 0.76$ for boys, and $\lambda = 0.90$ for girls. We get that boys are more self-confident than girls, on average but also for all levels of ability in line with Proposition 1.1b. The confidence level of boys increases with objective ability from 0.49 to 1, and the confidence level of girls increases from 0.35 to 1 in line with Proposition 1.1a. For an objective probability of success of 0.5, boys' subjective probability of success is equal to 0.86 and girls' subjective probability of success is equal to 0.79. In addition, boys' participation rate is equal to 61% and girls' participation rate is equal to 18%, in line with Proposition 2.1. The proportion of girls among the participants is 22.8%, and the participation gap is higher among the more able. Finally, the success rate for boys is 31% (more than three boys out of ten succeed among participating boys) while the success rate for girls is 9% (less than one girl out of ten succeeds among participating girls) in line with Proposition 4. The proportion of girls among the "winners" is 13.6%. All these results are summarized in Table 1.

Insert Table 1 here.

The table also contains results in the setting of Example 2Q with the specification (α, k, η, K) = (0.8, 1, 0.8, 0.1), and negative stereotype levels of $\lambda = 0.8$ for boys and $\lambda = 0.9$ for girls. We get very similar results as Example 1Q—girls have lower confidence, self-select less, and perform worse than boys.

Figures 1 and 2 represent the value of participation and self-confidence levels as a function of ability for boys and girls in Examples 1Q and 2Q. These figures show that the value of participation first decreases then increases with ability in line with Proposition 2.2, and that self-confidence levels weakly increase with ability in line with Proposition 1.1a.

Insert Figure 1 and 2 here.

4 Discussion and conclusion

Our model shows how differences in status can translate into differences in confidence, differences in choices and even differences in performance. In our approach, status imposes others' ability

[10] Assumptions (A1)-(A3) and (A5)-(A7) are satisfied without further restriction. Assumption (A4) is satisfied for $\lambda \geq 0.23$ and the condition on v_y is satisfied.

expectations and dictates the way failure and success will be interpreted, which translates into levels of self-confidence and choices through the individual's need to protect her self-esteem.[11] High status provides the individual's self-esteem with an insulating layer that enables more self-confidence and more bold choices without fearing the impact of failure in terms of self-esteem loss. Lower status individuals have a lower right to fail, hence will be more timid in their beliefs and choices leading to the maintenance of status hierarchies. Our model provides some new insight on a possible channel through which low status translates into more modest self-confidence levels and more timid choices.

An important application of our model is the issue of gender and math. There is evidence of a stereotype that girls have lower abilities when it comes to math.[12] There is also evidence of stereotyped attribution of success and failure in terms of ability,[13] so that Assumptions (A5) and (A6) are valid regarding the gender stereotype in mathematics. As a consequence of Proposition 1.1b, we obtain that girls have lower confidence levels in mathematics, consistent with evidence [see Correll, 2001, Buser et al., 2014]. Moreover, the positive correlation between self-confidence levels and ability in Proposition 1.1a is in line with educational statistics and surveys. For instance, in the Programme for International Student Assessment (PISA), science self-efficacy, which is a good proxy for self-confidence about academic ability in science, is positively correlated with the PISA score [Filippin and Paccagnella, 2012]. As far as choices are concerned, we obtain as a consequence of Proposition 1.2 and Proposition 2 that girls are underrepresented in math fields and that they choose less risky/difficult options in math, which is supported by evidence. In addition, Proposition 1.1b and Proposition 2.1 imply that participation in the risky option weakly increases with self-confidence, which is consistent with empirical evidence on math choices. For instance, Correll [2001] shows that controlling for objective ability, the higher students assess their mathematical ability, the greater the odds of enrolling in a high school calculus course and choosing a college major in science, math, or engineering. Finally, our prediction about the less positive relationship between self-selection and ability for girls in Proposition 4 and Corollary 3 is supported by the evidence in

[11] This is in line with what Ridgeway [2014, p. 5] notes about status beliefs: "Because individuals expect others to judge them according to these beliefs, they must take status beliefs into account in their own behavior, whether or not they personally endorse them."

[12] See Foschi [1996], Correll [2001], Rudman et al. [2001], Kiefer and Sekaquaptewa [2007], Reuben et al. [2014].

[13] For instance, for Yee and Eccles [1988], mothers think that talent is a more important explanation for boys' math successes while effort is a more important explanation for girls' math successes. See, also, Betz and Hackett [1981], Eccles and Jacobs [1986], Dweck [1986], Jacobs and Weisz [1994], Tiedemann [2000], Dickhäuser and Meyer [2006].

Ellison and Swanson [2010]. Ellison and Swanson (2010) show that the gender gap in math contests is even more pronounced than in the SAT, especially at the right tail of the ability distribution, suggesting a self-selection bias.

Besides gender and math, Assumptions (A5) and (A6) also apply more generally to gender and male stereotyped fields and more generally to gender issues. Indeed, there is a stereotype of lower ability of women in many fields [Foschi, 1996, Fiske et al., 2002] and there is evidence that women tend to take less credit for success and to blame their failures on lack of ability, especially in male stereotyped fields [Beyer, 1990]. Our analysis predicts lower self-confidence and lower participation among women in more ambitious/risky options. This is consistent with empirical evidence on women's confidence and on their lower representation at the top of the labor market hierarchy. Besides, the predictions of our model are consistent with experimental results on competitiveness—in particular, the lower self-selection of women as well as the less positive relation between self-selection and ability [Niederle and Vesterlund, 2007, Grosse and Reiner, 2010, Niederle et al., 2013, Buser et al., 2017].

Finally, our approach can be applied to any negatively stereotyped group, as long as the stereotype is salient and Assumptions (A5) and (A6) are satisfied. For instance, there is evidence that students from low socio-economic status (SES) attribute success more externally than high SES students [Mooney and Thornton, 1999]. Our model predicts their lower self-confidence as well as their lower participation in selective tracks or prestigious fields, controlling for ability. These two features are consistent with empirical evidence.[14] More generally, concerning educational or occupational issues, the mechanisms highlighted in this paper provide possible (partial) reasons why orientation choices and inequalities get perpetuated across generations.

We have shown that, without intervention, stereotypes or status hierarchies survive and even get stronger. Let us illustrate, in the setting of gender and math, which type of policy interventions might be efficient in light of the predictions of our model. Such interventions should not only aim to get *more* girls to participate, but they should especially target the most able girls, thereby reducing the gender self-selection bias and leading to increased performance of girls and a reduction

[14] For example, OECD [2013, Chapter 7, p. 91] notes that "Disadvantaged students are generally less likely to feel confident about their ability to tackle specific mathematics tasks than advantaged students. While these differences partly reflect differences in mathematics performance related to socio-economic status, these differences remain large and statistically significant even when comparing students who perform similarly in mathematics." See also Dar and Getz [2007].

of the stereotype of lower ability. Overall, our model suggests that girls should be provided with a safety cushion to their self-esteem in order to reduce the impact of stereotypes. The intervention that seems the most appropriate according to our analysis is the creation of mentor programs or of benevolent advice. Knowing that others expect one to do well at mathematics is precisely what boys enjoy and what girls lack, and what is at the origin of the differences in self-confidence and choices. As does a positive stereotype for boys, mentor programs/benevolent advice would provide negatively stereotyped groups with an insulating layer from the self-esteem threat that any achievement-related choice represents.

A Proofs

All proofs assume that $W''(y) < 0$ resulting from Assumptions (A1) and (A3), and each proof mentions which other specific hypotheses in Assumptions (A1)-(A6) are needed.

Proof of Proposition 1. 1a. We have either $\frac{\partial y^*}{\partial p} = 0$ or $\frac{\partial y^*}{\partial p} = -\frac{\varphi_y(x_h,y^*,\lambda)-\varphi_y(x_l,y^*,\lambda)}{v_{yy}(y^*)+p\varphi_{yy}(x_h,y^*,\lambda)+(1-p)\varphi_{yy}(x_l,y^*,\lambda)}$, in which case $\frac{\partial y^*}{\partial p}$ has the sign of $\varphi_y(x_h, y^*, \lambda) - \varphi_y(x_l, y^*, \lambda)$, which is non-negative according to (A4).

1b. We have $\frac{\partial y^*}{\partial \lambda} = -\frac{p\varphi_{y\lambda}(x_h,y^*,\lambda)+(1-p)\varphi_{y\lambda}(x_l,y^*,\lambda)}{v_{yy}(y^*)+p\varphi_{yy}(x_h,y^*,\lambda)+(1-p)\varphi_{yy}(x_l,y^*,\lambda)}$, hence $\frac{\partial y^*}{\partial \lambda}$ has the sign of $p\varphi_{y\lambda}(x_h, y^*, \lambda) + (1-p)\varphi_{y\lambda}(x_l, y^*, \lambda)$, which is non-positive if and only if $\frac{p[\varphi_{y\lambda}(x_h,y^*,\lambda)-\varphi_{y\lambda}(x_l,y^*,\lambda)]}{-\varphi_{y\lambda}(x_l,y^*,\lambda)} \leq 1$. For $\frac{p[\varphi_{y\lambda}(x_h,y^*,\lambda)-\varphi_{y\lambda}(x_l,y^*,\lambda)]}{-\varphi_{y\lambda}(x_l,y^*,\lambda)} \geq 1$, and under

$$v_y(x_h) \geq \sup_{\lambda} \frac{\varphi_{y\lambda}(x_h, y^*, \lambda)[-\varphi_y(x_l, x_h, \lambda)] + (-\varphi_{y\lambda}(x_l, y^*, \lambda))[-\varphi_y(x_h, x_h, \lambda)]}{\varphi_{y\lambda}(x_h, y^*, \lambda) - \varphi_{y\lambda}(x_l, y^*, \lambda)}, \tag{A.1}$$

(also referred to as $v_y(x_h)$ large enough), we have $y^* = x_h$ and $\frac{\partial y^*}{\partial \lambda} = 0$; indeed, under (A.1), we have $\varphi_y(x_h, x_h, \lambda) - \varphi_y(x_l, x_h, \lambda) \geq [-\varphi_y(x_l, x_h, \lambda) - v_y(x_h)] \frac{[\varphi_{y\lambda}(x_h,y^*,\lambda)-\varphi_{y\lambda}(x_l,y^*,\lambda)]}{-\varphi_{y\lambda}(x_l,y^*,\lambda)}$ hence if $\varphi_y(x_l, x_h, \lambda) + v_y(x_h) \leq 0$, we have $p[\varphi_y(x_h, x_h, \lambda) - \varphi_y(x_l, x_h, \lambda)] \geq [-\varphi_y(x_l, x_h, \lambda) - v_y(x_h)] \frac{p[\varphi_{y\lambda}(x_h,y^*,\lambda)-\varphi_{y\lambda}(x_l,y^*,\lambda)]}{-\varphi_{y\lambda}(x_l,y^*,\lambda)} \geq [-\varphi_y(x_l, x_h, \lambda) - v_y(x_h)]$ and we know by the first order condition that this leads to $y^* = x_h$. Notice that if $\varphi_y(x_l, x_h, \lambda) + v_y(x_h) > 0$, the first order condition implies that $y^* = x_h$ for all p and we also obtain $\frac{\partial y^*}{\partial \lambda} = 0$.

2. We have $W(y^*) = v(y^*) + pu(x_h) + p\varphi(x_h, y^*, \lambda) + (1-p)\varphi(x_l, y^*, \lambda)$, hence $\frac{\partial W(y^*)}{\partial p} = v_y(y^*)\frac{\partial y^*}{\partial p} + u(x_h) + \varphi(x_h, y^*, \lambda) - \varphi(x_l, y^*, \lambda) + p\varphi_y(x_h, y^*, \lambda)\frac{\partial y^*}{\partial p} + (1-p)\varphi_y(x_l, y^*, \lambda)\frac{\partial y^*}{\partial p}$, with

$v_y(y^*)+p\varphi_y(x_h,y^*,\lambda)+(1-p)\varphi_y(x_l,y^*,\lambda)=0$ or $\frac{\partial y^*}{\partial p}=0$, hence $\frac{\partial W(y^*)}{\partial p}=u(x_h)+\varphi(x_h,y^*,\lambda)-\varphi(x_l,y^*,\lambda)>0$. Besides, we have $\frac{\partial W(y^*)}{\partial\lambda}=v_y(y^*)\frac{\partial y^*}{\partial\lambda}+p\varphi_y(x_h,y^*,\lambda)\frac{\partial y^*}{\partial\lambda}+(1-p)\varphi_y(x_l,y^*,\lambda)\frac{\partial y^*}{\partial\lambda}+p\varphi_\lambda(x_h,y^*,\lambda)+(1-p)\varphi_\lambda(x_l,y^*,\lambda)$ with $v_y(y^*)+p\varphi_y(x_h,y^*,\lambda)+(1-p)\varphi_y(x_l,y^*,\lambda)=0$ or $\frac{\partial y^*}{\partial\lambda}=0$, hence $\frac{\partial W(y^*)}{\partial\lambda}=p\varphi_\lambda(x_h,y^*,\lambda)+(1-p)\varphi_\lambda(x_l,y^*,\lambda)\le 0$, due to Assumption (A5). ■

Proof of Proposition 2. Participation is characterized by $V(p,\lambda)=pu(x_h)-(k+1)u(A(p))+v(y^*)+p\varphi(x_h,y^*,\lambda)+(1-p)\varphi(x_l,y^*,\lambda)\ge 0$.

1. We have $V_\lambda(p,\lambda)=v_y(y^*)\frac{\partial y^*}{\partial\lambda}+p\varphi_y(x_h,y^*,\lambda)\frac{\partial y^*}{\partial\lambda}+(1-p)\varphi_y(x_l,y^*,\lambda)\frac{\partial y^*}{\partial\lambda}+p\varphi_\lambda(x_h,y^*,\lambda)+(1-p)\varphi_\lambda(x_l,y^*,\lambda)$ with either $v_y(y^*)+p\varphi_y(x_h,y^*,\lambda)+(1-p)\varphi_y(x_l,y^*,\lambda)=0$ or $\frac{\partial y^*}{\partial\lambda}=0$, hence $V_\lambda(p,\lambda)=p\varphi_\lambda(x_h,y^*,\lambda)+(1-p)\varphi_\lambda(x_l,y^*,\lambda)\le 0$ due to Assumption (A5).

2. We have $V_p(p,\lambda)=u(x_h)-(k+1)\frac{\partial u(A(p))}{\partial p}+v_y(y^*)\frac{\partial y^*}{\partial p}+\varphi(x_h,y^*,\lambda)-\varphi(x_l,y^*,\lambda)+p\varphi_y(x_h,y^*,\lambda)\frac{\partial y^*}{\partial p}+(1-p)\varphi_y(x_l,y^*,\lambda)\frac{\partial y^*}{\partial p}$, with either $v_y(y^*)+p\varphi_y(x_h,y^*,\lambda)+(1-p)\varphi_y(x_l,y^*,\lambda)=0$ or $\frac{\partial y^*}{\partial p}=0$, hence $V_p(p,\lambda)=u(x_h)-(k+1)\frac{\partial u(A(p))}{\partial p}+\varphi(x_h,y^*,\lambda)-\varphi(x_l,y^*,\lambda)$. We have $V_{pp}(p,\lambda)=-(k+1)\frac{\partial^2 u(A(p))}{\partial^2 p}+[\varphi_y(x_h,y^*,\lambda)-\varphi_y(x_l,y^*,\lambda)]\frac{\partial y^*}{\partial p}$. We know that $\frac{\partial^2 u(A(p))}{\partial^2 p}\le 0$ (by Assumption (B)), that $\varphi_y(x_h,y^*,\lambda)-\varphi_y(x_l,y^*,\lambda)\ge 0$ by Assumption (A4), and that $\frac{\partial y^*}{\partial p}\ge 0$ by Proposition 1. The function $V(\cdot,\lambda)$ is then convex hence its section $\{p:V(p,\lambda)\le 0\}$ is convex and participation is characterized by $p\in[0,p_0(\lambda)]\cup[p_1(\lambda),1]$. The function V_p is increasing in p. If $V_p(x_l,\lambda)$ is positive then V_p is positive and participation is weakly increasing in p. If $V_p(x_h,\lambda)$ is negative then V_p is negative, and participation is weakly decreasing in p. If $V_p(x_l,\lambda)$ is non-positive and $V_p(x_h,\lambda)$ non-negative, then the function V_p is first non-positive, and non-negative above a given threshold $\widehat{p}_\lambda$ characterized by $V_p(\widehat{p}_\lambda)=0$. The value of participation and participation are weakly decreasing on $[0,\widehat{p}_\lambda[$ then weakly increasing on $]\widehat{p}_\lambda,1]$. ■

Proof of Corollary 3. 1. As in the proof of Proposition 2, and adopting the same notations, we have $V_p(p,\lambda)=u(x_h)-(k+1)\frac{\partial u(A(p))}{\partial p}+\varphi(x_h,y^*,\lambda)-\varphi(x_l,y^*,\lambda)$.

When $V=0$, we have $pu(x_h)-(k+1)u(A(p))+v(y^*)+p\varphi(x_h,y^*,\lambda)=-(1-p)\varphi(x_l,y^*,\lambda)$, hence $(1-p)V_p(p,\lambda)|_{V=0}=u(x_h)-(1-p)(k+1)\frac{\partial u(A(p))}{\partial p}-(k+1)u(A(p))+v(y^*)+\varphi(x_h,y^*,\lambda)$, whose derivative with respect to λ is equal to $B=v_y(y^*)\frac{\partial y^*}{\partial\lambda}+\varphi_y(x_h,y^*,\lambda)\frac{\partial y^*}{\partial\lambda}+\varphi_\lambda(x_h,y^*,\lambda)$. Now, $B=[-p\varphi_y(x_h,y^*,\lambda)-(1-p)\varphi_y(x_l,y^*,\lambda)]\frac{\partial y^*}{\partial\lambda}+\varphi_y(x_h,y^*,\lambda)\frac{\partial y^*}{\partial\lambda}+\varphi_\lambda(x_h,y^*,\lambda)=(1-p)[\varphi_y(x_h,y^*,\lambda)-\varphi_y(x_l,y^*,\lambda)]\frac{\partial y^*}{\partial\lambda}+\varphi_\lambda(x_h,y^*,\lambda)$. We have $\varphi_y(x_h,y^*,\lambda)-\varphi_y(x_l,y^*,\lambda)\ge 0$ by Assumption (A4), $\varphi_\lambda(x_h,y^*,\lambda)\le 0$ by Assumption (A5), and $\frac{\partial y^*}{\partial\lambda}\le 0$ under the condition given in

(A.1) on v_y, hence $B \leq 0$.

2a. We know by the proof of Proposition 2 that the function V is convex in p. It suffices to show that $V_p(p,\lambda)|_{V=0} \leq 0$ to get that participation is weakly decreasing with ability. As seen in the proof of Corollary 3.1, we have $(1-p)V_p(p,\lambda)|_{V=0} = u(x_h) - (1-p)(k+1)\frac{\partial u(A(p))}{\partial p} - (k+1)u(A(p)) + v(y^*) + \varphi(x_h, y^*, \lambda)$. Since $u(A(p))$ is concave by assumption, we have $\frac{\partial u(A(p))}{\partial p} \geq \frac{u(x_h)-u(A(p))}{1-p}$ hence $(1-p)V_p(p,\lambda)|_{V=0} \leq -v(x_h) + \varphi(x_h, y^*, \lambda) + v(y^*)$. For $\lambda = 1$, we have $\varphi(x_h, y^*, 1) = 0$ and $(1-p)V_p(p,\lambda)|_{V=0} \leq 0$ for all $y^* \leq x_h$ due to (A1).

2b. As in the proof of Proposition 2 we know that participation is weakly increasing if $V_p(x_l,\lambda)$ is positive, and in particular, if $\varphi(x_h, x_l, \lambda)$ is high enough. ■

Proof of Proposition 4. 1. For $\lambda_1 = 1$, $\varphi(x_h, y, 1) = 0$, hence according to Corollary 3, participation of group G_1 is then given by $[0, p_0(1)]$. By Proposition 2, we know that participation of group G_2 is given by $p \in [0, p_0(\lambda_2)] \cup [p_1(\lambda_2), 1]$, and since participation weakly decreases with λ (Proposition 2), with $p_0(\lambda_2) \geq p_0(1)$. We compare the expected values conditional on participation $E^{G_1}[\widetilde{p}] = \frac{E\left[\widetilde{p}1_{[0,p_0(1)]}(\widetilde{p})\right]}{E\left[1_{[0,p_0(1)]}(\widetilde{p})\right]}$ and $E^{G_2}[\widetilde{p}] = \frac{E\left[\widetilde{p}1_{[0,p_0(\lambda_2)]\cup[p_1(\lambda_2),1]}(\widetilde{p})\right]}{E\left[1_{[0,p_0(\lambda_2)]\cup[p_1(\lambda_2),1]}(\widetilde{p})\right]}$. Since $1_{[0,p_0(1)]}(\widetilde{p}) = 1_{[0,p_0(1)]}(\widetilde{p})\, 1_{[0,p_0(\lambda_2)]\cup[p_1(\lambda_2),1]}(\widetilde{p})$ and $1_{[0,p_0(1)]}(\widetilde{p})$ is a weakly decreasing function of $\widetilde{p}$, we then have $E^{G_1}[\widetilde{p}] \leq E^{G_2}[\widetilde{p}]$ for $\lambda_1 = 1$. Indeed, we have $E^{G_2}\left[\widetilde{p}1_{[0,p_0(1)]}(\widetilde{p})\right] \leq E^{G_2}[\widetilde{p}]\; E^{G_2}\left[1_{[0,p_0(1)]}(\widetilde{p})\right]$ hence $\frac{E^{G_2}\left[\widetilde{p}1_{[0,p_0(1)]}(\widetilde{p})\right]}{E^{G_2}\left[1_{[0,p_0(1)]}(\widetilde{p})\right]} \leq E^{G_2}[\widetilde{p}]$ and $\frac{E^{G_2}\left[\widetilde{p}1_{[0,p_0(1)]}(\widetilde{p})\right]}{E^{G_2}\left[1_{[0,p_0(1)]}(\widetilde{p})\right]} = E^{G_1}[\widetilde{p}]$. Note that if $P(\widetilde{x} \in [p_1(\lambda_2), 1]) > 0$, then the previous inequality is strict and by continuity, also holds for λ_1 high enough. If $P(\widetilde{x} \in [p_1(\lambda_2), 1]) = 0$, then $E^{G_2}[\widetilde{p}] = \frac{E\left[\widetilde{p}1_{[0,p_0(\lambda_2)]}(\widetilde{p})\right]}{E\left[1_{[0,p_0(\lambda_2)]}(\widetilde{p})\right]}$ and $E^{G_1}[\widetilde{p}] \leq E^{G_2}[\widetilde{p}]$ for all $\lambda_1 \in [\lambda_2, 1]$.

2. In the setting of Bell [1985]'s model, as shown in Proposition A.1, participation for G_1 is given by $p \in [0, p_0(\lambda_1)]$ and participation for G_2 is given by $[0, p_0(\lambda_2)]$ with $p_0(\lambda_2) \geq p_0(\lambda_1)$. We have $E^{G_i}[\widetilde{p}] = \frac{E\left[\widetilde{p}1_{[0,p_0(\lambda_i)]}(\widetilde{p})\right]}{E\left[1_{[0,p_0(\lambda_i)]}(\widetilde{p})\right]}$. Since $1_{[0,p_0(\lambda_1)]}(\widetilde{p}) = 1_{[0,p_0(\lambda_1)]}(\widetilde{p})\, 1_{[0,p_0(\lambda_2)]}(\widetilde{p})$ and $1_{[0,p_0(\lambda_1)]}(\widetilde{p})$ is a weakly decreasing function of $\widetilde{p}$, we then have $E^{G_1}[\widetilde{p}] \leq E^{G_2}[\widetilde{p}]$. ■

Proposition A.1 *In the setting of* ***Example 1Q:***

1. *The optimal self-confidence level is given by $y^* = 0$ for $p \leq \frac{\lambda\eta - k}{\lambda\eta - (1-\lambda)K}$, by $y^* = 1$ for $p \geq \frac{\lambda\eta - k(1-\alpha)}{\lambda\eta - (1-\lambda)K}$ and by $y^* = \frac{k - p(1-\lambda)K - \lambda\eta(1-p)}{k\alpha}$ otherwise.*

2. *The optimal self-confidence level y^* is weakly increasing in p if $\lambda \geq \frac{K}{(\eta+K)}$, and weakly decreasing in λ if $k(1-\alpha) \geq \frac{\eta K}{(\eta+K)}$; it is weakly increasing in k, weakly decreasing in η, and weakly decreasing in K.*

3. *Participation is characterized for* $p \leq \frac{\lambda\eta-k}{\lambda\eta-(1-\lambda)K}$ *by* $(1-\lambda)K \geq k\left(1-\frac{1}{2}\alpha\right)$, *for* $p \geq \frac{\lambda\eta-k(1-\alpha)}{\lambda\eta-(1-\lambda)K}$ *by* $k\left(1-\frac{1}{2}\alpha\right) \geq \lambda\eta$ *and for* $\frac{\lambda\eta-k}{\lambda\eta-(1-\lambda)K} \leq p \leq \frac{\lambda\eta-k(1-\alpha)}{\lambda\eta-(1-\lambda)K}$ *by* $\frac{1}{2}\left(k-p(1-\lambda)K - \lambda\eta(1-p)\right)^2 - k\alpha p\left[k\left(1-\frac{1}{2}\alpha\right)-K(1-\lambda)\right] \geq 0.$

4. *Participation is characterized by* $p \in [0, p_0(\lambda)]$ *for some* $p_0(\lambda) \in [0,1]$.

5. *If* $K \leq \left(1-\frac{1}{2}\alpha\right)$, *then controlling for confidence, participation still weakly decreases with the intensity of the stereotype* λ.

Proof. 1. Immediate. 2. Condition on v_y in (A.1) is given by $k(1-\alpha) \geq \frac{\eta K}{(\eta+K)}$. The rest is immediate. 3. For $p \leq \frac{\lambda\eta-k}{\lambda\eta-(1-\lambda)K}$, $y^* = 0$, and participation is characterized by $K(1-\lambda) \geq k\left(1-\frac{1}{2}\alpha\right)$. For $p \geq \frac{\lambda\eta-k(1-\alpha)}{\lambda\eta-(1-\lambda)K}$, $y^* = 1$, and participation is characterized by $k\left(1-\frac{1}{2}\alpha\right) \geq \lambda\eta$. Otherwise, participation is characterized by $k\left(y-\frac{1}{2}\alpha y^2\right) - kp\left(1-\frac{1}{2}\alpha\right) + p(1-\lambda)K(1-y) - (1-p)\lambda\eta y \geq 0$. Direct computations lead to $\frac{1}{2}k\alpha y^2 - p\left[k\left(1-\frac{1}{2}\alpha\right)-K(1-\lambda)\right] \geq 0$ or $\frac{1}{2}\left(k-p(1-\lambda)K-\lambda\eta(1-p)\right)^2 - k\alpha p\left[k\left(1-\frac{1}{2}\alpha\right)-K(1-\lambda)\right] \geq 0$.

4 a. If $\lambda\eta > k\left(1-\frac{1}{2}\alpha\right)$, then, as seen in 3., participation is never chosen for $p \geq \frac{\lambda\eta-k(1-\alpha)}{\lambda\eta-(1-\lambda)K}$, hence according to Proposition 2, given by $p \leq p_0(\lambda)$.

b. If $\lambda\eta \leq k\left(1-\frac{1}{2}\alpha\right)$, then let us show that $S = [0,1]$. We have $k-p(1-\lambda)K-\lambda\eta(1-p) \geq k-p(1-\lambda)K-k\left(1-\frac{1}{2}\alpha\right)(1-p) = p\left[k\left(1-\frac{1}{2}\alpha\right)-(1-\lambda)K\right]+\frac{1}{2}k\alpha$, hence $\frac{1}{2}\left(k-p(1-\lambda)K-\lambda\eta(1-p)\right)^2 \geq \frac{1}{2}\left(p\left[k\left(1-\frac{1}{2}\alpha\right)-(1-\lambda)K\right]+\frac{1}{2}k\alpha\right)^2 \geq 2p\left[k\left(1-\frac{1}{2}\alpha\right)-(1-\lambda)K\right]\left(\frac{1}{2}k\alpha\right) = k\alpha p\left[k\left(1-\frac{1}{2}\alpha\right)-(1-\lambda)K\right]$.

5. Consider two levels of the stereotype λ_1 and λ_2 associated with the same level of confidence y^*. For $y^* = 0$, we have $V = -pv(1)+p(1-\lambda_i)K = p\left[-v(1)+(1-\lambda_i)K\right]$ hence $V < 0$ if $v(1) > K$. For $y^* = 1$, we have $V = (1-p)\left[v(1)-\lambda\eta\right]$, hence V has the sign of $v(1)-\lambda\eta$ and participation weakly decreases with λ. Otherwise, consider two individuals denoted by 1 and 2 with $y_1^* = y_2^* = y$ for $(\lambda_1, p_1) \geq (\lambda_2, p_2)$ and letting $V^{(i)} = v(y) - p_i v(1) + p_i\varphi(1, y, \lambda_i) + (1-p_i)\varphi(0, y, \lambda_i)$, let us show that $V^{(1)} \leq V^{(2)}$. We have $V^{(1)}-V^{(2)} = (p_2-p_1)v(1)+\left[p_1(1-\lambda_1)-p_2(1-\lambda_2)\right]K(1-y)+\left[(1-p_2)\lambda_2-(1-p_1)\lambda_1\right]\eta y$. Since $y_1^* = y_2^*$, we have $\lambda_2\eta(1-p_2)-\lambda_1\eta(1-p_1) = p_1(1-\lambda_1)K - p_2(1-\lambda_2)K$, hence $V^{(1)}-V^{(2)} = (p_2-p_1)v(1)+\left[p_1(1-\lambda_1)-p_2(1-\lambda_2)\right]K = (p_2-p_1)\left[v(1)-K\right] + \left[-p_1\lambda_1+p_2\lambda_2\right]K \leq 0$. ■

Proposition A.2 *In the setting of* ***Example 2Q:***

1. *The optimal self-confidence level is given by* $y^* = 0$ *for* $p \leq \frac{\lambda\eta - k}{\lambda\eta}$, *by* $y^* = 1$ *for* $p \geq \frac{\lambda\eta - k(1-\alpha)}{\lambda\eta}$ *and by* $y^* = \frac{k - \lambda\eta(1-p)}{k\alpha}$ *otherwise.*

2. *The optimal self-confidence level* y^* *is weakly increasing in* p *and weakly decreasing in* λ*; it is weakly increasing in* k*, weakly decreasing in* η*, independent from* K*.*

3. *Participation is characterized for* $p \leq \frac{\lambda\eta - k}{\lambda\eta}$ *by* $K(1-\lambda) \geq k\left(1 - \frac{1}{2}\alpha\right)$*, for* $p \geq \frac{\lambda\eta - k(1-\alpha)}{\lambda\eta}$ *by* $K(1-\lambda)p \geq \left(\lambda\eta - k\left(1 - \frac{1}{2}\alpha\right)\right)(1-p)$ *and for* $\frac{\lambda\eta - k}{\lambda\eta} \leq p \leq \frac{\lambda\eta - k(1-\alpha)}{\lambda\eta}$ *by* $\frac{1}{2}(k - \lambda\eta(1-p))^2 - k\alpha p\left[k\left(1 - \frac{1}{2}\alpha\right) - K(1-\lambda)\right] \geq 0$.

4. *Participation is characterized by* $p \in [0, p_0(\lambda)] \cup [p_1(\lambda), 1]$ *for some* $p_0(\lambda)$ *and* $p_1(\lambda)$ *in* $[0, 1]$.

Proof of Proposition A.2. Immediate, proceeding as in the proof of Proposition A.1. ■

References

G. A. Akerlof and W. T. Dickens. The economic consequences of cognitive dissonance. *American Economic Review*, 72:307–319, 1982.

D. E. Bell. Disappointment in decision making under uncertainty. *Operations Research*, 33:1–27, 1985.

J. Berger, B. Cohen, and M. J. Zeldich. Status characteristics and social interaction. *American Sociological Review*, 37(3):241–255, 1972.

N. Betz and G. Hackett. The relationship of career-related self-efficacy expectations to perceived career options in college women and men. *Journal of Counseling Psychology*, 28:399–410, 1981.

S. Beyer. Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology*, 59(5):960–970, 1990.

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

M. K. Brunnermeier and J. A. Parker. Optimal expectations. *American Economic Review*, 95: 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.

T. Buser, N. Peter, and S. C. Wolter. Gender, willingness to compete and career choices along the whole ability distribution. DISCUSSION PAPER SERIES 10976, Institute of Labor Economics, 2017.

A. Caplin and J. Leahy. Psychological expected utility theory and anticipatory feelings. *The Quarterly Journal of Economics*, 116:55–79, 2001.

S. Correll. Gender and the career choice process: the role of biased self-assessments. *American Journal of Sociology*, 106:1691–1730, 2001.

Y. Dar and S. Getz. Learning ability, socioeconomic status, and student placement for undergraduate studies in Israel. *Higher Education*, 2007.

O. Dickhäuser and W.-U. Meyer. Gender differences in young children's math ability attributions. *Psychology Science*, 48(1):3–16, 2006.

E. Diener, C. Colvin, W. Pavot, and A. Allman. The psychic costs of intense positive affect. *Journal of Personality and Social Psychology*, 61(3):492–503, 1991.

C. S. Dweck. Motivational processes affecting learning. *American Psychologist*, 41(10):1040–1048, 1986.

J. S. Eccles and J. E. Jacobs. Social forces shape math attitudes and performance. *Signs*, 11(2): 367–380, 1986.

G. Ellison and A. Swanson. The gender gap in secondary school mathematics at high achievement levels: evidence from the American mathematics competitions. *Journal of Economic Perspectives*, 24(2):109–128, 2010.

L. Festinger. *A theory of cognitive dissonance*. Stanford University Press, 1957.

A. Filippin and M. Paccagnella. Family background, self-confidence and economic outcomes. *Economics of Education Review*, 31:824–834, 2012.

S. Fiske, A. Cuddy, P. Glick, and J. Xu. A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82:878–902, 2002.

M. Foschi. Double standards in the evaluation of men and women. *Social Psychology Quarterly*, 59:237–254, 1996.

M. Foschi. Double standards for competence: theory and research. *Annual Review of Sociology*, 26 (1):21–42, 2000.

C. Gollier and A. Muermann. Optimal choice and beliefs with ex-ante savoring and ex-post disappointment. *Management Science*, 56:1272–1284, 2010.

N. Grosse and G. Reiner. Explaining gender differences in competitiveness: gender-task stereotypes. Working Paper Jena University, 2010.

L. Guiso, F. Monte, P. Sapienza, and L. Zingales. Culture, gender, and math. *Science*, 320: 1164–1165, 2008.

F. Gul. A theory of disappointment aversion. *Econometrica*, 59:667–686, 1991.

J. E. Jacobs and V. Weisz. Gender stereotypes: implications for gifted education. *Roeper Review*, 16(3):152–155, 1994.

W. James. *Principles of Psychology.* Harvard University Press, 1890.

J. S. Joensen and H. S. Nielsen. Mathematics and gender: heterogeneity in causes and consequences. *The Economic Journal*, 126(593):1129–1163, 2016.

E. Jouini, P. Karehnke, and C. Napp. On portfolio choice with savoring and disappointment. *Management Science*, 60(3):796–804, 2014.

N. Karlsson, G. Loewenstein, and J. W. Patty. A dynamic model of optimism. Working Paper Carnegie Mellon University, 2004.

B. Kőszegi and M. Rabin. A model of reference-dependent preferences. *The Quarterly Journal of Economics*, 121:1133–1165, 2006.

B. Kőszegi. Ego utility, overconfidence and task choice. *Journal of the European Economic Association*, 4:673–707, 2006.

A. K. Kiefer and D. Sekaquaptewa. Implicit stereotypes, gender identification, and math-related outcomes: a prospective study of female college students. *Psychological Science*, 18(1):13–18, 2007.

L. J. Kirkeboen, E. Leuven, and M. Mogstad. Field of study, earnings, and self-selection. *The Quarterly Journal of Economics*, 131(3):1057–1111, 2016.

W. Kopczuk and J. Slemrod. Denial of death and economic behavior. *Advances in Theoretical Economics*, 5(1):5, 2005.

G. Loewenstein. Anticipation and the valuation of delayed consumption. *The Economic Journal*, 97:666–684, 1987.

G. Loewenstein and P. Linville. Expectation formation and the timing of outcomes: a cognitive strategy for balancing the conflicting incentives for savoring success and avoiding disappointment. Working Paper, 1986.

G. Loomes and R. Sugden. Disappointment and dynamic consistency in choice under uncertainty. *Review of Economic Studies*, 53(2):271–282, 1986.

R. Macera. Dynamic beliefs. *Games and Economic Behavior*, 87:1–18, 2014.

L. Mechtenberg. Cheap talk in the classroom: how biased grading at school explains gender differences in achievements, career choices and wages. *The Review of Economic Studies*, 76(4): 1431–1459, 2009.

B. Mellers and A. McGraw. Anticipated emotions as guides to choice. *Current Directions in Psychological Science*, 10:210–214, 2001.

B. Mellers, A. Schwartz, K. Ho, and I. Ritov. Decision affect theory: emotional reactions to the outcomes of risky operations. *Psychological Science*, 8:423–429, 1997.

E. S. Mooney and C. A. Thornton. Mathematics attribution differences by ethnicity and socioeconomic status. *Journal of Education for Students Placed at Risk*, 4(3):321–332, 1999.

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

M. Niederle, C. Segal, and L. Vesterlund. How costly is diversity? Affirmative action in light of gender differences in competitiveness. *Management Science*, 59:1–16, 2013.

M. Niederle and L. Vesterlund. Explaining the gender gap in math test scores: the role of competition. *Journal of Economic Perspectives*, 24(2):129–44, 2010.

OECD. Pisa 2012 results: ready to learn (volume III). Technical report, OECD Publishing, Paris, 2013.

E. Reuben, P. Sapienza, and L. Zingales. How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences*, 111(12):4403–4408, 2014.

C. L. Ridgeway. Why status matters for inequality. *American Sociological Review*, 79(1):1–16, 2014.

C. Ridgeway. Gender, status, and leadership. *Journal of Social Issues*, 57:637–655, 2001.

L. A. Rudman, A. G. Greenwald, and D. E. McGhee. Implicit self-concept and evaluative implicit gender stereotypes: self and ingroup share desirable traits. *Personality and Social Psychology Bulletin*, 27(9):1164–1178, 2001.

J. Tiedemann. Parents' gender stereotypes and teachers beliefs as predictors of children's concept of their mathematical ability in elementary school. *Journal of Educational Psychology*, 92:144–151, 2000.

B. Weinberg. A model of overconfidence. *Pacific Economic Review*, 14:502–515, 2009.

C. Weinberger. Is the science and engineering workforce drawn from the far upper tail of the distribution? Working Paper University of California Santa Barbara, 2005.

D. K. Yee and J. S. Eccles. Parent perceptions and attributions for children's math achievement. *Sex Roles*, 19(5-6):317–333, 1988.

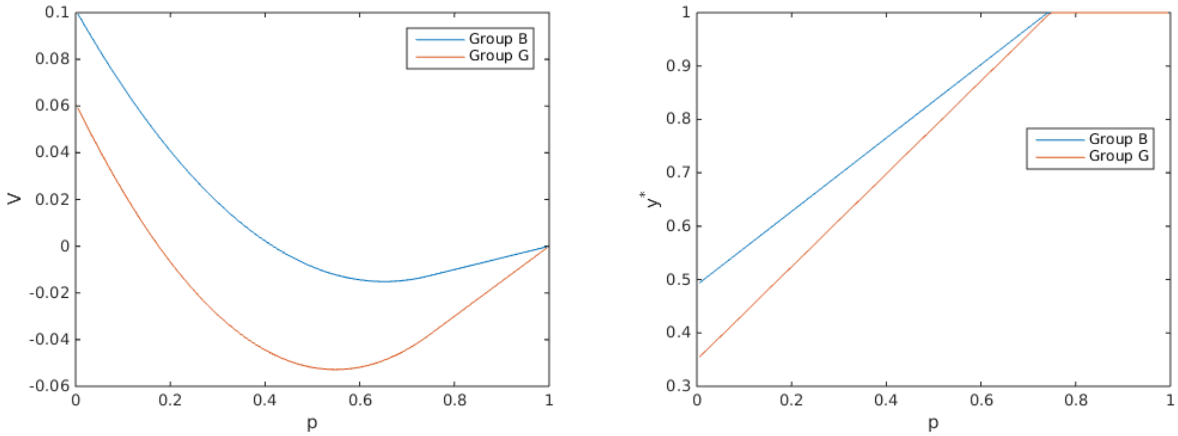

Figure 1: The left-hand side figure shows the value of participation as a function of ability and the right-hand side figure shows the self-confidence level as a function of ability. Both figures are for Example 1Q and the specification $(\alpha,\, k,\, \eta,\, K) = (0.8,\, 1.25,\, 1,\, 0.3)$ and $\lambda = 0.76$ for Group B (boys) and $\lambda = 0.9$ for Group G (girls).

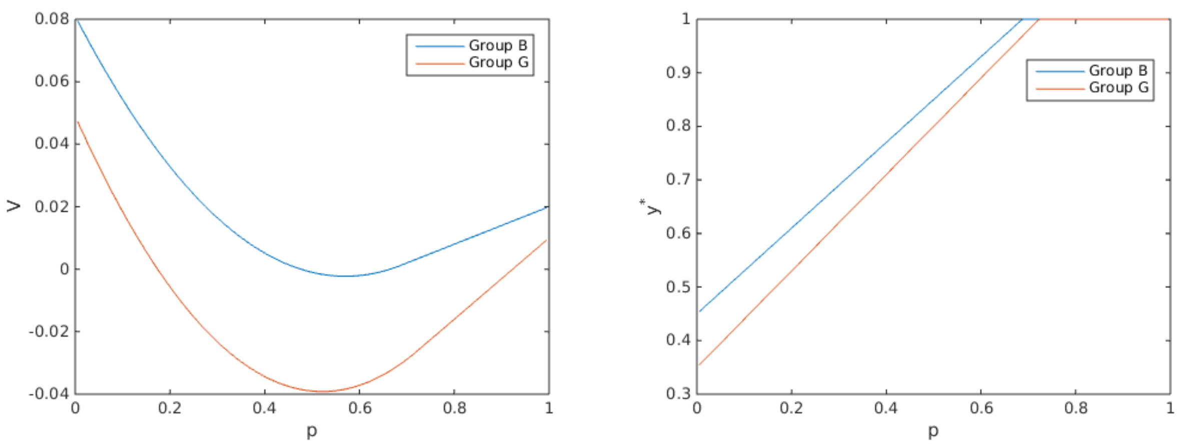

Figure 2: The left-hand side figure shows the value of participation as a function of ability and the right-hand side figure shows the self-confidence level as a function of ability. Both figures are for Example 2Q and the specification $(\alpha,\, k,\, \eta,\, K) = (0.8,\, 1,\, 0.8,\, 0.1)$ and $\lambda = 0.8$ for Group B (boys) and $\lambda = 0.9$ for Group G (girls).

Table 1: The table contains numerical results for Example 1Q with the specification (α, k, η, K) = (0.8, 1.25, 1, 0.3) and for Example 2Q with the specification (α, k, η, K) = (0.8, 1, 0.8, 0.1). The probability of success follows a uniform distribution on $[0, 1]$.

	Example	Stereotype	Self-confidence y^*		Participation	Ability cond.
		λ	Interval	Average	for $A = CE$	on participation
Group B	1	0.76	[0.49, 1]	0.81	61%	0.31
Group G	1	0.90	[0.35, 1]	0.76	18%	0.09
Group B	2	0.80	[0.45, 1]	0.81	81%	0.48
Group G	2	0.90	[0.35, 1]	0.77	25%	0.35