PARIS SCHOOL OF ECONOMICS
ECOLE D'ECONOMIE DE PARIS

ENS
ÉCOLE NORMALE SUPÉRIEURE

PSL RESEARCH UNIVERSITY PARIS

ANALYSIS AND POLICY IN ECONOMICS

PARIS SCHOOL OF ECONOMICS

Gender norms and rational choices of women in STEM: a study of the ENS' entrance exams

September 7th, 2022

Supervisor: Thomas Breda

Referee: Eric Maurin

Clémence Gardette

Gender norms and rational choices of women in STEM: a study of the ENS' entrance exams

Clémence Gardette
under the supervision of
Thomas Breda

September 2022

Abstract

The under-representation of women in STEM has been intensely studied by researchers and policymakers, since attaining an equal representation of women in science is part of a global policy promoting gender equality. However, recent studies seem to show that women tend to avoid STEM topics not because of discrimination, but because of their personal preferences (themselves shaped by gender norms). We are using a particular empirical setting, the entrance exams of the ENS (Ecole Normale Supérieure de Paris), to measure whether gender norms still exist when individuals make a choice in a context where that choice only impact a binary outcome (entering the school) and nothing else, and as a consequence where the decision-making should be closer to rational models of decision. We show that gender norms impact the topics chosen by students in the more gender-segregated track, but that this is not the case in the less gender-segregated track. This is consistent with our theoretical model assuming that the strength of a gender norms is the share of women in a track. This model also shows that, under certain circumstances, it could be rational for the school to apply positive discrimination to correct for the gendered choices made by students, as was shown in Breda and Ly, 2015. Finally, we exploit a survey distributed to candidates in order to better understand their decision-making, and whether they are aware of gender norms associated with scientific topics.

I would like to thank Thomas Breda for his helpful advice and guidance as my supervisor, as well as Eric Maurin for accepting to be the referee. The ENS was incredibly helpful both for the access to the data and for the organization of the survey; in particular, Ronald Caruana and Emanuel Basset provided invaluable support. I am deeply grateful for Ryan Lahfa for his support and his help with R.

JEL codes: I23, J16, D80

Contents

1 Introduction

It is a well-asserted fact that women tend to study STEM[1] less than men; the proportion of women studying mathematical-intensive topics is always lower than the proportion of women in higher education, and this is documented in all countries. Many public policies try to remediate this fact, in particular by encouraging women to choose STEM majors (in both high-schools and higher education) and by helping them to gain confidence in their mathematical skills (see for instance Calvez and Viry, 2018, two French MPs who wrote an information report on the situation of women in science in France, containing recommendations in order to improve the share of women in STEM). These policies are part of a more global movement of equality between men and women; in fact, the under-representation of women in STEM careers contributes to the observed pay gap between men and women (since STEM workers benefit from higher average wages), and some asserted that the under-representation of women in STEM induces inefficiencies.

Many authors have studied the reasons explaining the lower share of women in STEM. While discrimination against women and lower performance of women in mathematical-intensive may play some role, most authors generally agree on the fact that the main factor is the preferences of women themselves. Of course, these preferences are shaped by social interactions and stereotypes; however, in most contexts, it is impossible to distinguish between a choice that is guided by gender norms and a choice that is merely the consequence of personal preferences. For instance, a young woman who choose to major in Humanities in university may choose to do so despite her high ability in mathematics, and implementing public policies that respect the freedom of choice while correcting for possibly internalized stereotypes is impossible in most contexts without reliable information available to disentangle the two phenomena.

Our paper uses a particular empirical setting, the entrance exams of the Ecole Normale Supérieure de Paris (ENS), in which candidates have to make a choice that will have an impact on their results at the exams, but not the major they will choose later on. In this context, personal preferences between candidates play no role; the only goal of the choice is to maximize the probability to enter the school. As a consequence, it allows us to measure whether social norms play a role even when no preference is involved. Three tools are used in order to answer this question. Firstly, we wrote a theoretical model which offers a justification for making irrational choices when taking into account the gender norms. This theoretical model also provides a possible reaction from the school side. Secondly, we used administrative data given by the ENS in order to estimate whether candidates choices are influenced by social norms. Finally, we submitted a multiple-choice survey to this year's candidates in order to better understand the constitutive factors of these choices.

Three main conclusions can be derived from this work. First, we see that gendered norms are having an impact on individual choices, even when preferences are not involved; this effect seems to be stronger in more gender segregated tracks. Second, this individual effect fades out when considering the overall entrance exams: since men and women are both impacted by gendered choices, the share of women that enter the school is not different when suppressing the effect of the choice between candidates. Thirdly, it may, in this context, be rational for the school to implement some positive discrimination to account for the gendered choices (such positive discrimination was identified by Breda and Ly, 2015).

The rest of this paper is organized as follows. Section 2 presents the ENS entrance exams in details, as long as the literature to which this paper tries to contribute. Section 3 presents the three sources of data that are used. Section 4 presents the theoretical model of the option choice, and section 5 presents the methodology implemented. Finally, section 6 presents the results, and section 7 concludes.

[1]science, technology, engineering and mathematics

2 Background and Literature review

2.1 Background

The ENS entrance exams The École Normale Supérieure de Paris is one among many Grandes Ecoles, which exists in France aside the traditional universities, delivering a graduate level of diplomas (5 years of studies after high school). These Grandes Ecoles are very diverse, and include engineering schools, business schools, veterinarian and agronomic schools, and the four Écoles Normales Supérieures (ENS), which present the particularities to deliver a researched-oriented teaching in a wide variety of disciplines - for instance, the ENS of Paris, on which we will be focusing, allows its students to specialize in 15 different fields, from Mathematics to Philosophy. In order to get into the ENS of Paris, students have two main choices (we are not developing the tracks that recruit only a few people per year) : they can either study three years as undergraduate in any university, and then ask for an admission into the ENS' department corresponding to their specialization. We will call this track the "university track", and will not be focusing on it in this work; a little less than a half of the students, enter the ENS through this track. Students can also, right after high school, enter a CPGE, which are highly intensive two (or three) years preparatory schools that train the students to take the entrance exams of both the ENS and other Grandes Écoles. We will call this track the "CPGE track", and this track makes up for the half of the admissions in the ENS.

For the CPGE track, the entrance exam is divided in two parts : the written part ("admissibilité") and the oral track ("admission"). The written part takes place in April, is open to any candidate who choose to take it, and the inscription to the exam is free. The written exam consist of a few (between 4 and 6) written exams, focused on the topics of the tracks; these exams are quite long (between 4 and 6 hours) and difficult (the eligibility rate is less than 10% after written exams). After the written exams, the students who ranked above a certain score ("eligible students") are invited to the oral exams in June-July; these exams are focused on the same topics as the written exams, and take around one hour each. After this, students are ranked accordingly to the complete set of their weighted grades in their contest track, then, if they are above a certain score ("barre d'admission"), they are divided into a main ranking and secondary ranking, main ranking are guaranteed to enter to ENS, secondary ranking is a waiting list for the withdrawals from the main ranking list. Students who choose to enter the ENS will then begin a four-years education in September.

Science entrance tracks The CPGE track is itself divided in two categories, Science and Humanities, further divided according to the specialization of students. The science CPGE track entrance exams includes 5 different specializations, which correspond (with one exception, detailed below) to the type of CPGE student followed for two years : MP (Mathematics-Physics), Info (Computer science), PSI (Physics-Engineering sciences), PC (Physics-Chemistry) and BCPST (Biology-Geology). Since this thesis is focusing on the three tracks where students have a choice of option, the descriptive statistics of the sucess at exams are only reported for these three tracks (MP, PC and BCPST), in Table 1. The eligible candidates are the ones who succeeded at the written exams and the admitted candidates are the one who succeeded at the oral exams. However, these candidates are often admitted in other prestigious schools (mostly the Ecole Polytechnique, a military polivalent school), and some of them withdraw from the ENS. As a consequence, I also displayed the entered candidates, because this information is relevant in order to understand choices of women and because it of course impact the feminization of school.

Much information can be taken from Table 1. First, all the tracks are in fact highly selective : only a few dozens (at best) of the hundreds of candidates enter the ENS each year. Second, the proportion of women in each track is aligned with the proportion of women in higher education worldwide (Kahn

Table 1: Descriptive statistics on the tracks, gender and options of the candidates to an ENS entrance exam, for the years 2016-2018

	Track								
	MP			PC			BCPST		
Option	All	MP	MPI	All	Phy	Che	All	Bio	Geo
Number of candidates	6614	5334	1280	1470	859	611	3260	2686	377
Candidates per year (average)	1654	1334	320	368	215	153	815	672	94
Prop. of women candidate	18%	19%	12%	24%	16%	35%	59%	60%	54%
Eligible candidates per year (average)	143	100	44	103	65	38	73	62	11
Prop. of women eligible	6%	7%	3%	17%	12%	26%	53%	53%	50%
Admitted candidates per year (average)	38	26	12	19	14	5	21	16	5
Prop. of women admitted	11.0%	14.0%	4.0%	17.0%	15.0%	25.0%	49.0%	49.0%	48.0%
Entered candidates per year (average)	30	20	10	14	9	5	16	14	3
Prop. of women entered	7.0%	8.0%	5.0%	19.0%	19.0%	19.0%	55.0%	52.0%	73.0%

Note: in 2018, no data was included on the candidates who chose to enter the ENS; the entrance statistics are hence computed on the years 2016, 2017 and 2019 only

and Ginther, 2017): less than 1 candidate in 5 is a women in the MP track, there is a little bit more women in the PC track, and there is a (small) majority of women in the BCPST track, where the main topics are biology and geology. If the choice of tracks can be purely based on preferences, the choice of option should be, on the opposite, based on a rational choice (maximizing one's grade and hence the probability to pass the entrance exam). However, we see that option choices are gendered inside each track, which suggests either that men and women score differently according to the topic, or that these choices are impacted to social norms. Moreover, we see that in each track, women score less than men at written exams: there are fewer women who are eligible after the written exams than women who took the exams in the first place, in proportion. This is also in line with the observation of (Ceci and Williams, 2010), who showed that women tend to score a little bit lower than men at the top of the distribution. However, we see that there is an evolution of the share of women after the oral exams : for the less feminized tracks (all MP tracks, and the PC track), the share of women admitted is higher than the share of women eligible, while for more feminized tracks, the opposite is true. These results are in line with Breda and Ly, 2015, which suggests a positive discrimination for women in male dominated fields and a positive discrimination for men in women dominated fields at the ENS entrance exams (for the years 2004-2009).

It is important to note that the choice of the CPGE track is made by the students at the end of high-school, according to their tastes, grades, and career objectives ; once this choice is made, there is no possibility for a student to change from one track to another, except in some rare exceptions (and even in this case, students often have to do the two years of CPGE once again). Since we will be focusing on the entrance exam, we will hence consider the choice of track made by one student as a given. Moreover, since each entrance track correspond to one CPGE track, and since the number of places open is fixed by track, students from different CPGE tracks (for instance MP and PC track) are never directly in competition. We will hence consider the different tracks are totally independent of each other.

Options Almost all the entrance tracks of the ENS are further divided between options. Both Science and Humanities tracks include options, but they are organized quite differently : I will hence first focus on the Science track, before moving to the Humanity track.

On one hand, in three of the five science tracks, students can choose between two options : the PC track allows choosing between Physics and Chemistry, the BCPST track between Biology and Geology, and the MP track between Mathematics-Physics-Computer science and Mathematics-Physics; there is no option choice in PSI and Info tracks. The impact of the option choice is mainly to modify the weight of the exams taken by students ; in some tracks, and for some exams, it also impacts which exams are taken. All the different weights, for both written and oral exams, are detailed in Table 2.

In both the BCPST track and the PC track, this option choice is made by students individually for each track, and the choice is made late in CPGE (around January, hence three months before the written exams). As a consequence, this choice does not change the syllabus that the students have to study : the main impact of this choice is to change the weights that are put on the grades in order to compute the ranking. Moreover, this choice impacts only the entrance exam of the ENS ; it does not impact any other entrance exam the students might be taking (often, engineering schools' entrance exams). Finally, this choice does not impact the latter studies of the students, either outside or inside the ENS : the diploma they get after finishing CPGE does not mention this option, and the ENS insists heavily on the fact that the students are totally free to study in any field after entering the school, regardless of their option choice[2].

On the other hand, the Humanity track, which is itself divided into AL (Classics) and BL (Social Sciences) also makes the students choose an option for the exam. However, this track work slightly differently as compared to the Science track : first, students can choose between as many as eight choices for their option, while students in science tracks have only a choice between two options. It is then likely that the factors of the choice in the Humanity tracks are more complex to study and to understand. Second, the impact of this option choice is very different : in science track, the option does not change the curriculum, and does not impact the further studies of students outside the ENS. However, in Humanity tracks, students have to choose their option at the end of their first year of CPGE, and their second year of CPGE's curriculum is impacted by that choice. Moreover, if these students do not enter the ENS (which is the case of the vast majority of them, since very few seats are available in the entrance exams), and choose to go to the university, they have to specialize in university in the same track as their option. This entails that the option choice for the humanity students has no reason to be only determined by the chances of one student to enter the ENS, which is one characteristic of the Science tracks we exploit; as a consequence, we will leave this track out of this paper. However, the study of Humanity tracks, with adaptions, may represent an extension of the current paper.

2.2 Literature review

Women in STEM This thesis takes place in the abundant literature that describes, analyzes, and explains the place of women in STEM fields and higher education. It is well documented that women are a minority in STEM curricula worldwide (Kahn and Ginther, 2017, J. Delaney and Devereux, 2021): in OECD countries, fewer than 1 in 3 engineering graduates and fewer than 1 in 5 computer science graduates are girls. This fact is also well documented in France: for instance, Chabanon and Jouvenceau,

[2]This is in fact one of the first information displayed on the website page dedicated to the entrance exam : "Spécialité : Les spécialités choisies par les candidats pour passer les épreuves ne préjugent en rien de leur choix d'études ultérieurs."

Table 2: Table of weights in different tracks

	Track					
	MP		PC		BCPST	
Option	MP	MPI	Physics	Chemistry	Biology	Geology
Written exams						
Physics	6	4				
Mathematics C	4					
Computer Science A		6				
Mathematics D	6	6				
Physics*			7	5		
Chemistry*			5	7		
Mathematics			5	5		
Biology					7	4
Geology					2	5
Physics					2	3
Chemistry					4	3
Oral exams						
Physics	25					
Computer science		25				
Mathematics I	30	30				
Mathematics II	15	15				
Physics			26	20		
Chemistry			20	26		
Physics lab work			12			
Chemistry lab work				12		
Mathematics			20	20		
Biology					25	17
Geology					12	20
Physics					16	16
Chemistry					16	16
Lab work					12	12
Mathematics					16	16
Literature	8	8	8	8	8	8
Foreign Language I	3	3	3	3	3	3
Foreign Language II	3	3	3	3	4	4
TIPE	8	8	8	8	15	15
Sum of written weights	16	16	17	17	15	15
Sum of oral weights	92	92	100	100	127	127
Sum of all weights	108	108	117	117	142	142

*Note: An * indicates when the exam is on the same subject, but when two different exams take place according to the option chosen. Reading : the mathematics entrance exam in the PC track has a weight of 5, which represents one-third (5/15) of the written grade which allows the candidate to take the oral exams, and around 4% (5/117) percents of the final grade after oral exams. The TIPE is an undergraduate research project.*

2022 demonstrate in an INSEE[3] note that women in high school are far less likely to choose scientific topics to study (26% of girls choose only scientific topics, against 36.5% of boys). These number are even more striking when one accounts for the fact that the majority of higher education students are women since a few decades (Becker et al., 2010, Diprete and Buchmann, 2006): 57% of young women will enter tertiary education for the first time before they turn 25 on average across OECD countries, compared to 45% of men (OECD, 2021). In France, a recent analysis focusing on the Grandes Ecoles (Bonneau et al., 2021) showed that while women represent 55% of all the higher education students, only 40% of the students of the Grandes Ecoles were women, a proportion that even drops to around 40% when focusing on engineering schools.

When looking more closely at the numbers, one can moreover establish that women in STEM suffers from a "double attrition". On the one hand, the higher the level of education in any STEM field, the lower the proportion of women (Shapiro and Sax, 2011). This longitudinal attrition begins as early as kindergarten (Ceci et al., 2014), and continues through primary school (Bouffard et al., 2006), middle and high school (Chabanon and Steinmetz, 2018, Card and Payne, 2017) and at each stage of higher education. It is important to note that the effects of gender attrition through time are cumulative : J. M. Delaney and Devereux, 2019 show that more than half of the gendered gap between STEM and non-STEM majors in college applications is explained by different majors chosen in high school (rather than by grades obtained). It even continues once women are in their careers, with a progression of women scientists that may be slowed as compared to men scientists (Xie and Shauman, 2003), even though literature on this topic is contrasted (Ceci et al., 2014).

On the other hand, at any given fixed level of education, the more mathematics-intensive the field of the study, the fewer the proportion of women will be (Kahn and Ginther, 2017). Women are highly concentrated in psychology, social sciences and biological sciences, but they are much fewer in computer sciences (20% of all US-delivered bachelor degrees in 2019), engineering (23%) and mathematics (42%) (Trapani and Hale, 2022). This transversal attrition creates a highly segregated higher education (Grenet et al., 2021) : in France, in 2017, 78% of the gender segregation of the higher education in France in caused by the fact that men and women study different topics (rather than being segregated between highly selective and lowly selective tracks). This segregation is even more apparent in more egalitarian countries, despite girls scoring similarly or higher than boys in science, which is called the gender-equality paradox (Stoet and Geary, 2018).

This lower proportion of women studying STEM also contributes to the wage gap between men and women, since STEM careers benefits from higher wages and many professional opportunities (Weinberger, 1999) ; in 2010, women held less than 25% of STEM jobs in the US. Many authors also suggests that this lack of women in STEM may slow down innovation processes (Beede et al., 2011).

Understanding women under-representation in STEM While assessing the under-representation of women in STEM, authors tried to understand the mechanisms of this under-representation. In a rather basic framework (inspired of the one used in Ceci and Williams, 2010), three main factors may explain this under-representation : lower ability of women in STEM, discrimination and preferences, that may themselves be shaped by the social context women evolve in. The implications of these factors are rather different in terms of public policy. If the women, in fact, perform less than their masculine counterparts in STEM, policymakers may want to increase the education effort in STEM towards women, but only if this lower achievement is not genetically induced. However, if the under-representation of women in STEM is the consequence of discrimination, policymakers may want to implement a variety of systems to avoid such injustice. Finally, the preferences of women themselves may be the cause of this under-representation, which then calls for cautious action : individuals are allowed to study and work in the

[3]the French National Institute of Statistics and Economic Studies

fields they favor, and it is then unclear why the implication of the policymaker may be needed. Finally, it is worthy to note that these three causes are not mutually exclusive, and can reinforce themselves (for instance, Carlana, 2019 and Carrell et al., 2010 show that women that have been exposed to teachers with stereotypes perform less well in mathematics: discrimination can hence induce lower ability) ; I will however discuss them separately for more clarity.

Firstly, one possible cause of under-representation of women is the fact that they may not be performing as well as their male colleagues. Lots of research has been focused on trying to understand whether a biological difference between men and women may explain differences in STEM performances (for instance, large meta-studies have shown that spatial ability is decisive in STEM performances, and men perform better in spatial recognition (as early as 5-6 months old) - see Wai et al., 2009). However, Ceci and Williams, 2009 shows that mean results in mathematics are undifferentiated between men and women. The only difference in mathematical scores is the higher dispersion of men of achievement (there are more men at top and bottom of the distribution). Moreover, studies (Marks, 2008) have shown that STEM achievement of men and women varies significantly across countries, which suggests that if there is a part of biological determinism, some other factors are at play, and that policy action is possible and effective. In addition to the objective ability of women in STEM, some other factors correlated with ability may play a role. Some studies have been focusing on the effect of the rank (i.e. comparative ability), while controlling for a given ability, and have shown that women tend to rank lower than men for the same ability, and that rank explains partly the choice to pursue a STEM major (J. Delaney and Devereux, 2021). Finally, another possible explanation for the gender-gap in STEM is the higher ability of women in non-STEM (humanities, literature...) topics, which is widely assessed. Since women have a comparative advantage in non-STEM topics, their withdrawal from STEM majors may be a consequence of them exploiting this comparative advantage rather than scoring worse than men in STEM (Breda and Napp, 2019).

Secondly, another possible cause of women under-representation in STEM is that they may be discriminated against. If the existence of discrimination is well-assessed in laboratory experiments (see Anderson et al., 2005 for a review of economical and psychological experiments showing that there exist discrimination in favor of individuals from dominant groups), the literature assessing discrimination against women in STEM is rather nuanced. If some studies show that such discrimination does exist (Bloodhart et al., 2020: women are outperforming as compared to men, but are still considered less able), the majority of the literature tends to conclude that there is strong evidence that discrimination against women in STEM is small if not null (Ceci and Williams, 2010, Ginther and Kahn, 2006), at least since the beginning of the 2000s. Some papers, both theoretical and empirical, even argue that women may benefit from a positive discrimination in STEM (Fryer, 2007, Breda and Ly, 2015). However, discrimination may also be internalized by women: under this hypothesis, discrimination may be impossible to observe empirically, because women adjust their choices and behaviors before being actually discriminated against. There, in fact, exist some evidence that women have internalized the stereotype stating that "mathematics are not for women" (Breda et al., 2020). This internalized discrimination has been studied in psychology under the concept of "stereotype threat" (see Spencer et al., 1999 for the origins of stereotype threat, then highly replicated in literature, for instance in Kiefer and Sekaquaptewa, 2007 and O'Brien and Crandall, 2003), even though the role that stereotype threat in STEM gender gap may have been overstated (see Stoet and Geary, 2012 for a discussion of the importance of stereotype threat for women in STEM).

Finally, one last cause of women under-representation in STEM may be their personal preferences ; in fact, assuming that women choose their major and careers to maximize their utility, an aversion of women for STEM may simply explain why they are under-represented. This last possibility relies on "identity economics", as it has been forged by Akerlof and Kranton, 2010 : identity shape the economic choices we make, and hence it is important to take it into account when studying economic facts. In this case, literature has been focusing on trying to understand the construction of preferences, that are at least partially shaped by the social environment people evolve in. For instance, Brenøe and Zölitz, 2020 have shown that women who have more female peers in high school are less likely to pursue a STEM major, and that the effects are long-lasting. However, one difficulty with explaining the gender-gap with preferences is that these preferences may be hardly distinguishable from stereotype threats internalized by women (see Cheryan et al., 2017 for a psychological model trying to explain horizontal segregation of women, i.e. between biology and mathematics). Another difficulty with assuming that the attrition of women in STEM is caused by their individual preferences is that it implies that individuals take rational decisions based on objective criteria (rational choice theory), but there is some evidence that women largely underestimate their ability in mathematics as compared to boys, once one control by the ability (Perez-Felkner et al., 2017). This underestimation of women of their own ability is, once again, indistinguishable from preferences in most empirical studies.

3 Data

Three sets of data are used in this paper : data containing the results of the entrance exams of the ENS for the years 2016-2019 (for all tracks), data describing the education followed by students after their entrance at the ENS (during the years 2010-2020) and data provided by a multiple-choice survey, answered by the candidates to the ENS in June 2022. The current section is dedicated to the description of these datasets.

3.1 Entrance exams data

The first source of data used in this thesis is provided by the ENS : these data sets contain all the administrative data collected by the ENS for all the candidates in all tracks : BCPST, PSI, MP-Info, PC, AL and BL, for the years 2016-2019 included (except for the PSI track, which was created in 2018). The data for the year 2020 was also transmitted by the ENS, but this year was excluded from the start, because the pandemic had a huge impact on the ENS exams this year (in particular, no oral exams took place this year, and the written exams took place in June instead of April). As explained previously in the background part, our study focuses on three particular scientific tracks that allow students between two options : MP, PC and BCPST tracks. In addition to containing a record of all eligible and admitted candidates (see Table 1), these data sets also include all the grades obtained by the candidates in all the exams taken.

Grades These datasets also include the grades obtained by all the candidates at the written tests, and the grades obtained by the eligible candidates at the oral exams. For both the PC and the BCPST track, all the candidates, despite their specialization, take written and oral exams in both options; the only impact of the choice of option is the modification of the weight coefficients applied to grades (see Table 2). As a consequence, it is possible to compare the distribution of grades for both options, by gender, for the PC track (Figure 2) and for the BCPST track (Figure 3).

For the PC track (Figure 2), the interpretation that one can get from these plots are mixed. Regarding written exams, one can note that, in line with Table 1, women score worse than men in both topics, regardless of their option. However, the gap is wider for physics, which is more mathematics-intensive, as compared to chemistry. However, the picture is much more complicated for oral exams. Firstly, women score worse than men in physics, which is in line with the written exam results, but they score better than men for chemistry. This trend reversal is even more surprising given the fact that the positive

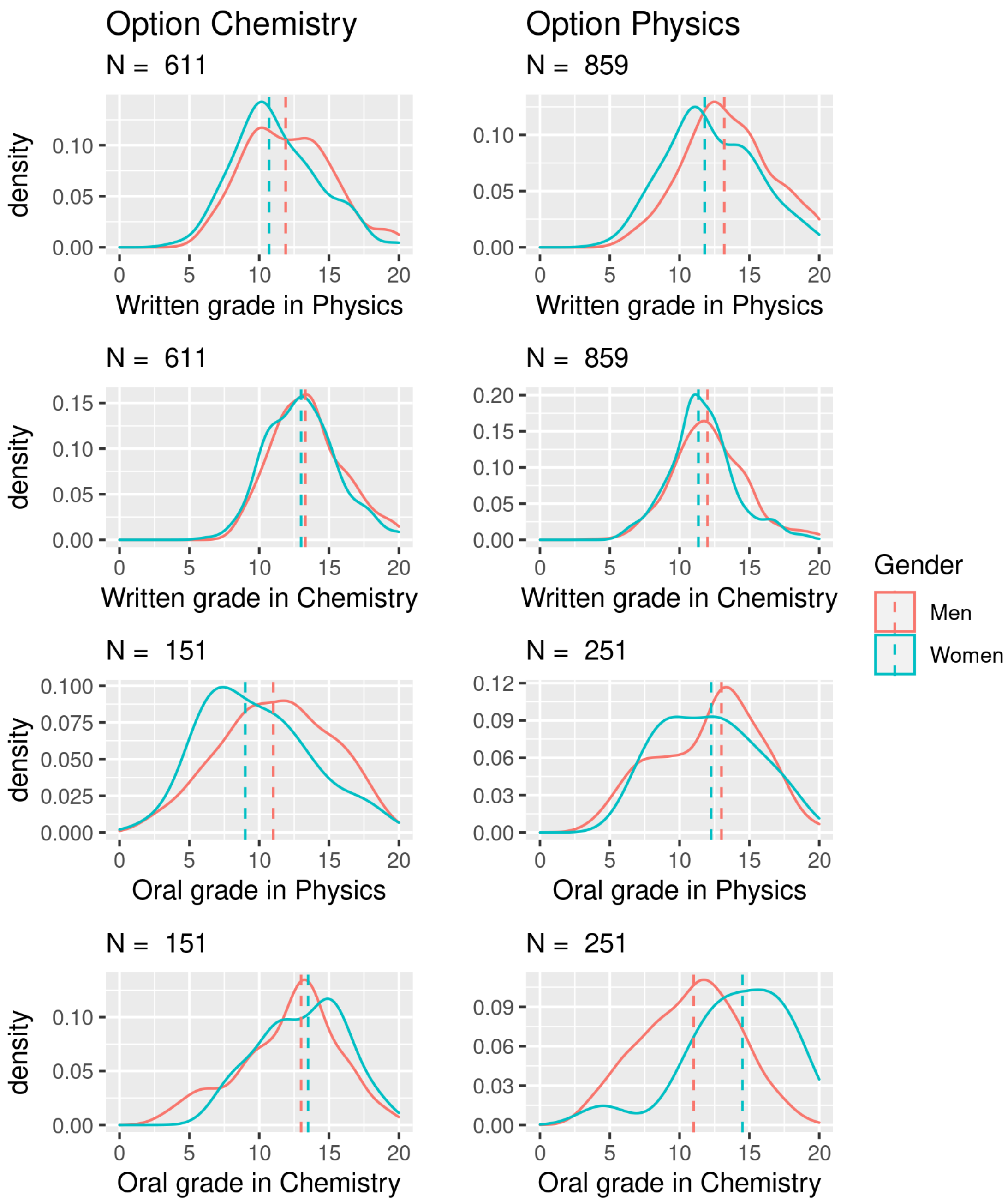


Figure 2: Distribution of grades in Physics and Chemistry in PC track, by option chosen

This figure plots the density of the distribution of the grades received by candidates in Physics and Chemistry, in the PC track. The left column depicts the density for the candidates who chose chemistry as an option, and the right column is for candidates who chose physics instead. The top two rows are plotting the written grades (for all candidates), and the bottom rows are plotting the oral grades (eligible candidates only). Finally, the dash lines represent the medians of the distribution, by gender.

discrimination in favor of women in more male-dominated fields would suggest the opposite - men should get a grade premium in chemistry, and women should get a premium in physics. Second of all, some choices of options made by students seems to be irrational. In fact, the median grade in chemistry of women candidates is higher for women who took the physics option than for women who took the chemistry option! Moreover, this effect goes the opposite way as compared to the stereotypes (according to which women should choose the chemistry option ”too much”).

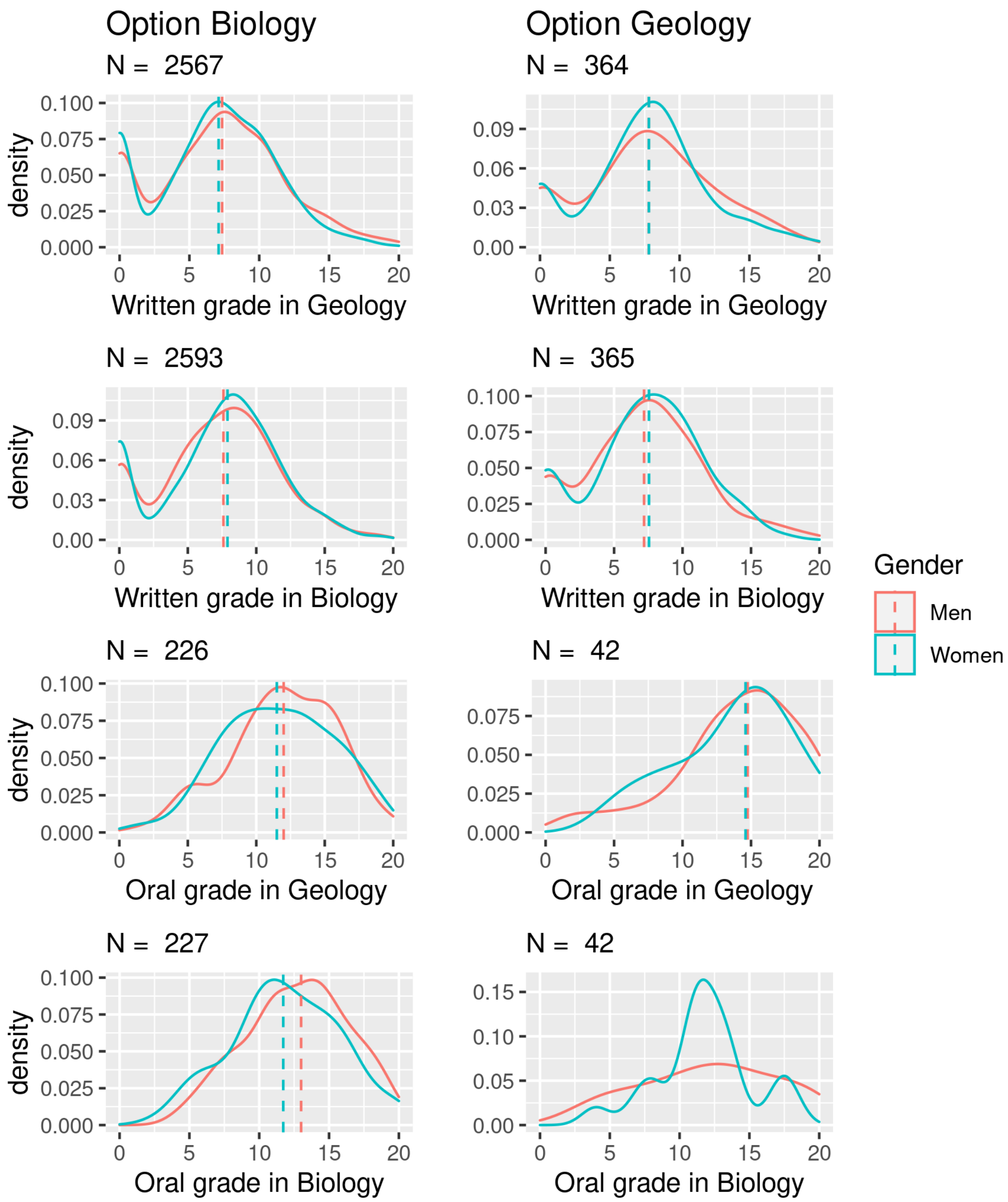


Figure 3: Distribution of grades in Biology and Geology in BCPST track, by option chosen

This figure plots the density of the distribution of the grades received by candidates in Biology and Geology, in the BCPST track. The left column depicts the density for the candidates who chose biology as an option, and the right column is for candidates who chose geology instead. The top two rows are plotting the written grades (for all candidates), and the bottom rows are plotting the oral grades (eligible candidates only). Finally, the dash lines represent the medians of the distribution, by gender.

For the BCPST track (Figure 3), the plots are much more straightforward to interpret. For all the written exams, the median results of both men and women are identical, and the distributions are very close. Moreover, we can see that there is almost no difference in the medians obtained with regard to the option chosen, which suggest that there is no strong differences in the average knowledge of individuals who choose different options. The picture is more contrasted for oral exams. We can see that candidates who choose the geology option tend to be a lot better than they are in biology (interestingly, the gap is higher for women than for men, suggesting that women have to be worse at biology to consider taking geology as an option), while the median grade in biology and in geology are closer for candidates who chose biology as an option. Biology seems to be the ”default” option, while geology is considered only by candidates who have a high achievement in geology; this is also correlated by the fact that only around 15% of candidates choose geology as an option, which seems a bit questioning.

The issue with the MP track In our background presentation of the ENS entrance exams (see section 2.1), we stated that three Science tracks offered an option to candidates: MP, PC and BCPST. However, while the PC and BCPST track offer a rather straightforward choice which modify only the weights of the exams taken, the MP track is much more complicated. First, the MP track is dedicated to the students who are in an MP CPGE; however, the Info track is also dedicated to the students who are in an MP CPGE. As a consequence, MP CPGE students have, in fact, three possible choices when taking the ENS entrance exam: MP-MP, MP-MPI, or Info. Second, as opposed to other tracks where the curriculum in CPGE is the same, regardless of the option chosen, MP students have to choose at the end of the first year of their CPGE between a computer science minor or an engineering science minor; they can only have one minor. Moreover, while the curriculum allows MP CPGE students with a computer science to take any of the three ENS entrance track, MP CPGE students with an engineering science minor can only take the MP-MP ENS exam (even though rare exceptions do exist). The data given by the ENS does not include the minor taken by the students in CPGE. Thirdly, as we can see in Table 2, even when considering the MP track independently of the Info one, we don't have any clear counterfactual to estimate the level in computer science of the MP, and the level in physics of the MPI, because the exams taken are different (the same problem arise when comparing the MPI and the Info tracks). As a consequence, our identification method, which rely on having the same students taking both an exam in her option and in a topic she did not choose as her option, fails; we will hence mainly exclude the MP track from our analysis, and focus on the PC and BCPST tracks. However, we distributed a survey to eligible candidates to understand their choice of options (section 3.3), and we will exploit some of the answers given by MP students in this context.

3.2 ENS' education data

The second source of data used in this thesis is also from an administrative source, and describes the education of ENS students once they entered the ENS. It is a panel dataset : for each year, there is one observation by student currently in the ENS. The department of enrollment in the ENS is included[4], for each year: in fact, students choose a department when they enter the school, and can change every year (in exchange for minor administrative frustrations). However, to graduate from the school, students have to fulfill a certain number of requirements relative to their departments: it is hence easier for students to change sooner (when they still have a few years before graduation), and it is not feasible for them to change too many times.

The data transmitted by the ENS includes all the students who studied in the ENS at least one year between 2016 and 2020, and their department. We first begin by restricting the dataset to students who entered between 2016 and 2019 (who are the ones for whom we have entrance exam data), and we then match them by their name and surname to the entrance exams table, for the three tracks we are interested in (MP, PC and BCPST). Four students were dropped (because there were no clear match between the two tables), leaving a total of 313 students. Their department are indicated in Table 4 (for the department of entry) and in Table 5 (for the department before graduating of the ENS, which is in

[4]There are 15 department in the ENS : 7 in science - neuroscience, computer science, biology, chemistry, geoscience, mathematics, physics - and 8 in humanities : arts, economics, history, philosophy, social sciences, ancient sciences, geography, and literature and languages

fact the last department recorded in our data, given that some students were not already in their last year of studies in 2020, which is the last year for which we have data).

Figure 4: Department of entry in the ENS, 2016-2019, by track of entry

			Department when entering the ENS							
Track	Option		Mat.	CSI.	Phy.	Che.	Bio.	Geo.	Neur. sc.	Hum.
PC	Che.	Total	2	1	8	15	0	1	0	0
		Women	0	0	1	3	0	0	0	0
	Phy.	Total	10	0	30	2	0	1	0	2
		Women	1	0	6	0	0	1	0	1
MP	MP	Total	88	0	11	0	0	0	0	0
		Women	7	0	2	0	0	0	0	0
	MPI	Total	47	8	0	0	0	0	0	0
		Women	3	0	0	0	0	0	0	0
BCPST	Bio.	Total	0	0	0	4	63	4	0	0
		Women	0	0	0	1	35	0	0	0
	Geo.	Total	0	0	0	0	6	8	0	0
		Women	0	0	0	0	4	5	0	0

Note: Mat. stands for mathematics, CSI for computer sciences, Phy. for physics, Che. for chemistry, Bio. for biology, Geo. for geology, Neur. sc. for neuroscience, Hum. for humanities (which includes all the Humanities department)

Figure 5: Department of graduation in the ENS, 2016-2019, by track of entry

			Department when graduating from ENS							
Track	Option		Mat.	CSI.	Phy.	Che.	Bio.	Geo.	Neur. sc.	Hum.
PC	Che.	Total	1	0	13	13	0	0	0	0
		Women	0	0	2	2	0	0	0	0
	Phy.	Total	9	0	32	0	0	2	0	2
		Women	1	0	5	0	0	2	0	1
MP	MP	Total	76	2	20	0	0	0	0	1
		Women	6	0	3	0	0	0	0	0
	MPI	Total	39	16	0	0	0	0	0	0
		Women	2	1	0	0	0	0	0	0
BCPST	Bio.	Total	0	0	0	4	63	3	1	0
		Women	0	0	0	1	34	0	1	0
	Geo.	Total	0	0	1	0	6	7	0	0
		Women	0	0	0	0	4	5	0	0

Note: Mat. stands for mathematics, CSI for computer sciences, Phy. for physics, Che. for chemistry, Bio. for biology, Geo. for geology, Neur. sc. for neuroscience, Hum. for humanities (which includes all the Humanities department)

These two tables show us that both the track and the option chosen inside each track are highly correlated with further choices of major in the ENS, despite the total freedom that students have after entering the ENS. For instance, for the PC track, only two students who took physics as an option go in chemistry, while they are 30 to pursue in physics; on the opposite, among the PC students who chose chemistry as an option, 15 of them pursue in chemistry, while 8 of them go in physics (for the entrance majors; the tendencies are the same with the department of graduation). The same pattern repeat in MP versus MPI, with a strong differentiation between physics and computer science, and in BCPST, between biology and geology. Moreover, these tables seems to indicate that in both PC and BCPST, the two options are not chosen in the same way, with both physics and biology being the "default option" - and the default major! It is striking that among the students who chose physics and biology, almost none choose to go in the "competing" department, while a significative proportion of the students who took the geology and of the chemistry options opt for the "attractive" (and in their case, not chosen as an option) department anyway. The choice of option, as the choice of department, are, in definitive, fairly asymmetrical between the two possibilities offered.

3.3 Multiple-choice survey

The last data set has been collected specifically for this paper, through a multiple-choice survey. The purpose of this survey was to understand better how candidates made their choices of option before taking the exam.

3.3.1 Target of the survey

In order to understand how the candidates made their choices of options, three subsets of people could have been the target of the survey: all the candidates, only the eligible candidates, or the admitted candidates. The ideal scenario would have been to be able to send the survey to all the candidates, before the written exam (or at least before the results of this exam), in order to keep survey answers unbiased by the results of the candidates in the exams. However, this was almost impossible to do, because there was no way to transmit the survey to all the candidates (email addresses were not provided). An alternative would have been to send the survey to some CPGE classes, but this would have implied a selection bias (choosing a "large" Parisian CPGE versus a small provincial one is likely to have an impact on the answers); we therefore didn't choose that approach. On the opposite, targeting all the admitted candidates would have been quite easy, because it is rather simple to share a survey on the social-media pages of ENS students. However, it would have been unclear how to interpret the results, because students currently in the ENS took the exam at least one year ago; their memories of their option choices are likely to be incorrect (in particular, some option choices can be justified or dismissed *a posteriori* after one receives her exam grades, or after she chooses her major). Moreover, this strategy would have implied to get answers only from students who succeeded at entering the ENS, another form of selection bias.

As a consequence, we chose to target the eligible candidates when conceiving and distributing the survey. Hence, we distributed the survey during the oral exams of the eligible candidates, for practical reasons (candidates are physically at the ENS for their exams) and because eligible candidates don't know their oral grades nor their written grades, which supports the idea that their answers will be unbiased. We received a strong support from the ENS, which allowed us to administer the survey in the room of *admisseurs*. This room is occupied by the administrative employees who are in charge of organizing all the practical organization of the oral exams (the *admisseurs*), and who are here every day to answer the practical questions of candidates (where to find their oral rooms, etc.). The survey was online, on the platform SurveyMonkey[5]. However, the candidates coming in the room of *admisseurs* were often only a few minutes from their oral exam, stressed and as a consequence not very interested in fulfilling an optional survey immediately. As a consequence, a flyer was given to the candidates, with a QR-Code redirecting towards the survey, and candidates were asked to fulfill it within a week. The flyer is reproduced in appendix A.1. Around 150 flyers were distributed in July 2022 (during around 10 days) for 71 answers collected. As a matter of fact, the answering rate is surprisingly high.

[5] https://fr.surveymonkey.com/

3.3.2 Contents of the survey

As stated before, the main goal of this survey was to understand the reasoning of the candidates when they chose their option: what has driven them to pick this particular choice? Were they consciously doing a rational choice (maximizing their probability of entering the ENS), or did other factors influenced their choices? The whole survey is reproduced in appendix A.2 (in French); this section summarizes and explain its content.

Questions 1 to 5: biographical information The first five questions of this survey are primarily descriptive; they ask the gender, the scholarship status, the number of times one took the exams, where they studied in CPGE, and their track (between MP, Info, BCPST, PC and PSI).

Questions 6 to 14: option chosen These questions mainly ask the candidates their option chosen, for each track (the candidate was only being shown the page corresponding to the track he or she chose in question 5; for instance, a PC candidate was directly asked to answer question 9). We also asked them the hour at which they took some of their oral exams.

Questions 15 to 22: core of the survey These eight questions are the core of the survey. They have two purposes: questions 15 to 20 try to understand how the candidates made their option choice, by asking them the factors of this choice, who helped them with this choice (teachers, family...), and if they regret this choice. Questions 20 to 22 try to measure if the gender stereotypes associated with topics are perceived by students, both consciously (we ask them if they feel like their option is more a "feminine" or "masculine" topic), and unconsciously (we ask them the gender of their option's teacher).

Questions 23 and 24: entering the ENS These questions try to measure whether the candidates understood that their option choice does not condition their major choice after entering the ENS.

Questions 25 to 30: material conditions of taking the exams These questions ask candidates whether they were in good material conditions (housing, transportation, etc.) when taking oral exams. They are orthogonal to this study, but were added in case the ENS was interested in the results (for instance to offer more support to disadvantaged students[6]).

Questions 31 to 33: merging the survey data with administrative data One goal when collecting this survey was to match its results with the administrative data collected by the ENS (entrance exam data and education data), in order to link these answers with actual results and to be able to test for certain hypothesis (for instance, whether individuals who evaluate topics as more stereotyped are more likely to score high on the topic corresponding to their gender and low on other topics). In order to do so, we then asked candidates their names, surnames and their explicit authorization for this matching. However, the survey were distributed in summer 2022, and the administrative data transmitted by the ENS did not include the results of the current year; we asked them the complementary data, but they weren't able to give it in time for this thesis (results for the exams are only complete at the beginning of August). As a consequence, the results of this survey were not matched with administrative data; this represents a possible extension of this research. However, the results of the survey can still be used as descriptive results, relevant for the average eligible candidate, and used to test hypotheses.

[6]The ENS already provides housing for some candidates; in addition, a regional scholarship as available to help students pay for the costs induced by the exams, but is not well-known.

3.3.3 Answers to the survey

As stated as the beginning of this subsection, 71 answers were collected for this survey. A document detailing all answers for every question of the survey is available in appendix A.3; since this document was produced for the ENS administration, it is written in French. This subsection is focused on some descriptive statistics given to understand the shares of the respondents in terms of sections and gender (corresponding to the questions 1 to 5 of the survey). The more psychological questions are interpreted and discussed in the Result section (section 6.4) of this paper.

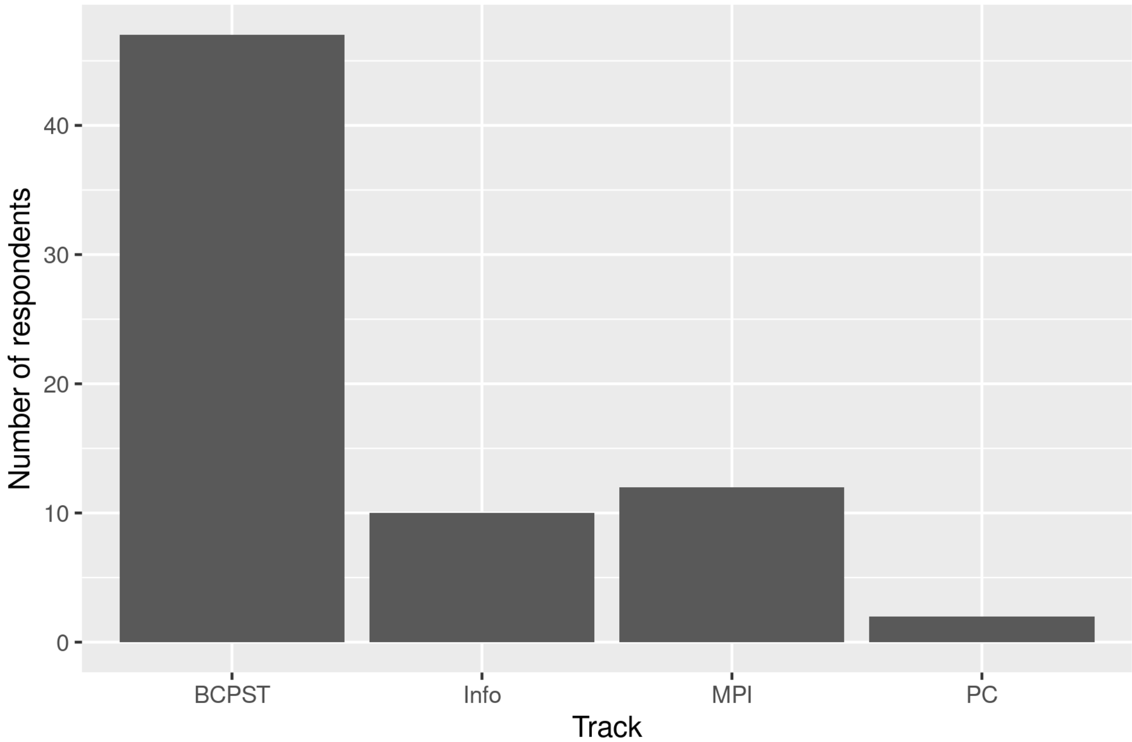


Figure 6: Distribution of the track of the respondents

Sections of the respondents As we see in Figure 6, there is an overwhelming majority of BCPST respondents to the survey (around 2/3 of all respondents); some answers were also collected in Info (a track where no option has to be chosen for the exams) and in MPI, though very few answers were collected in PC. This distribution can easily be explained by the practical organization of the exam. In fact, all BCPST candidates had to go to the room of *admisseurs* before their exams, which means that many of them received the flyer. On the opposite, Info and MPI candidates don't necessarily go in this room; they only meet the *admisseurs* when they encounter an issue, which means that fewer of them received the flyer in the first place. Finally, the very few ($N = 2$) PC respondents is explained by the fact that their exams take place in a different building, which is at a 5-minutes walk distance from the main building. As a consequence, even fewer are the PC candidates who go to the room of *admisseurs*, and hence who received the flyer. Our analysis of the survey will then mainly be focused on the BCPST respondents.

Gender of the respondents As we see in Figure 7, around two thirds of the respondents are men, and one-third are women. The picture is however quite different when plotting the gender of the respondents by track, as in Figure 8. In this case, we see that all the respondents in MPI and Info are men (and

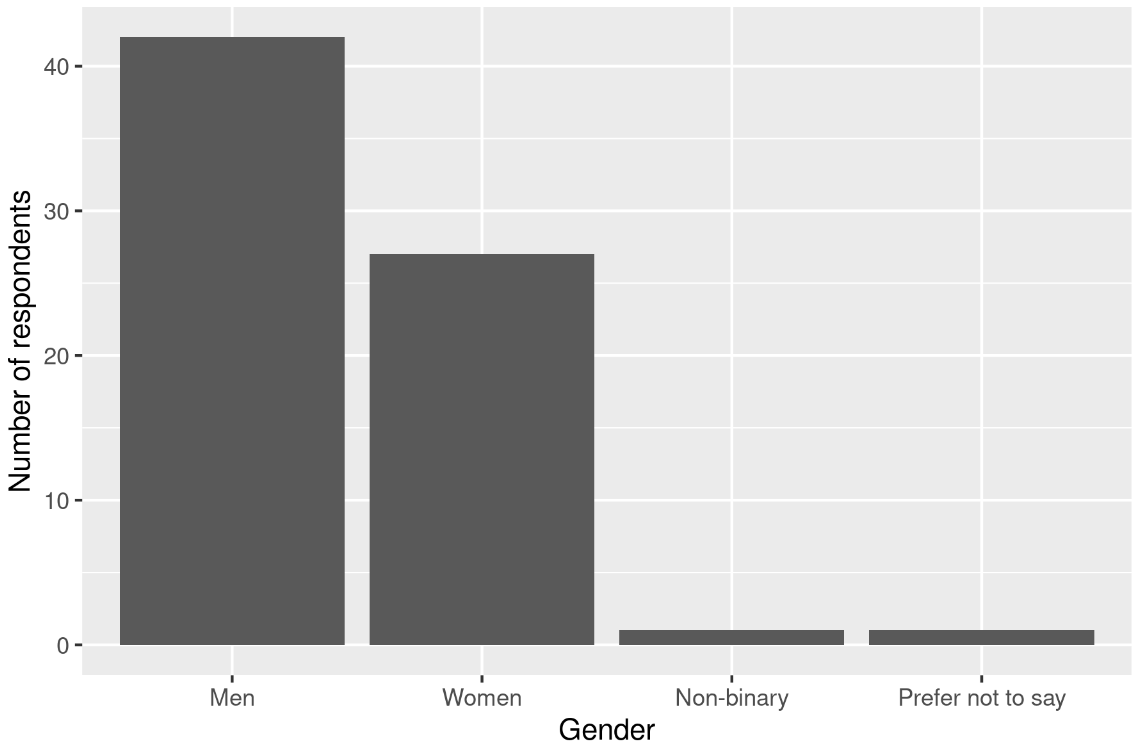


Figure 7: Distribution of the gender of the respondents

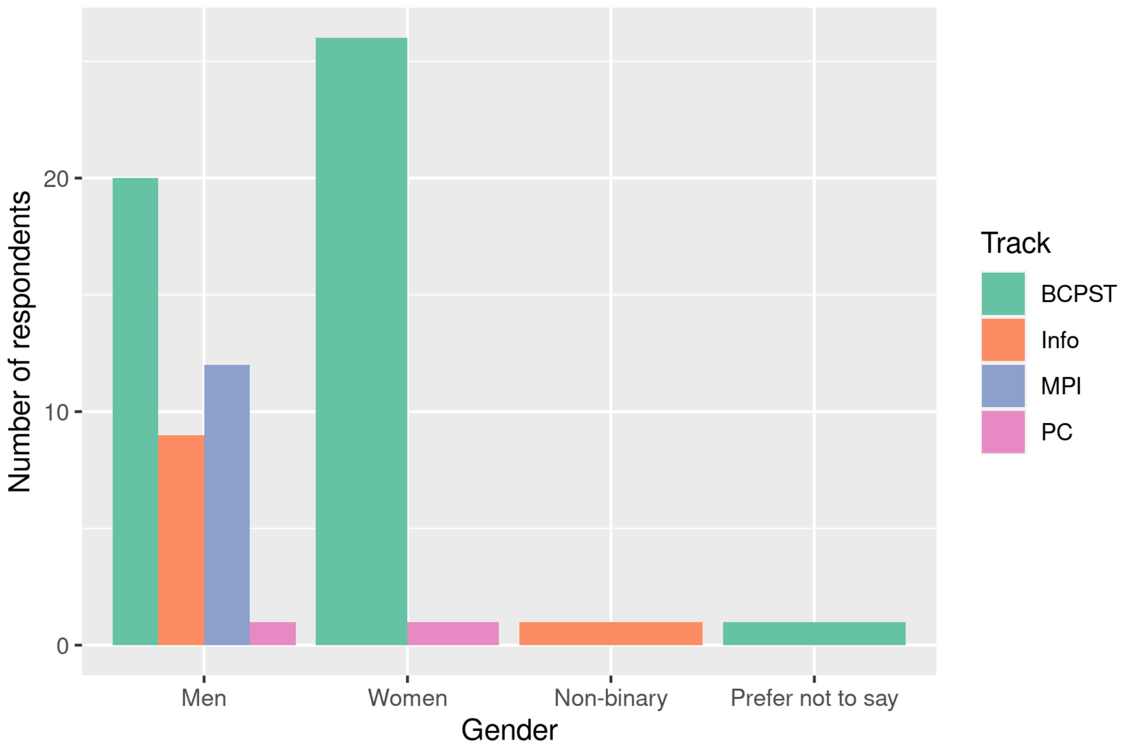


Figure 8: Distribution of the gender of the respondents, by track

one person who is non-binary). Since there is around 5% of women eligible in each of these tracks (see Table 1), it is coherent that no women appear in our sample. On the opposite, 57% of the respondents in BCPST are women, which is in line with the average proportion of eligible women in this track (Table

1). Due to the very small sample of answers in PC, no conclusion can be drawn from these statistics.

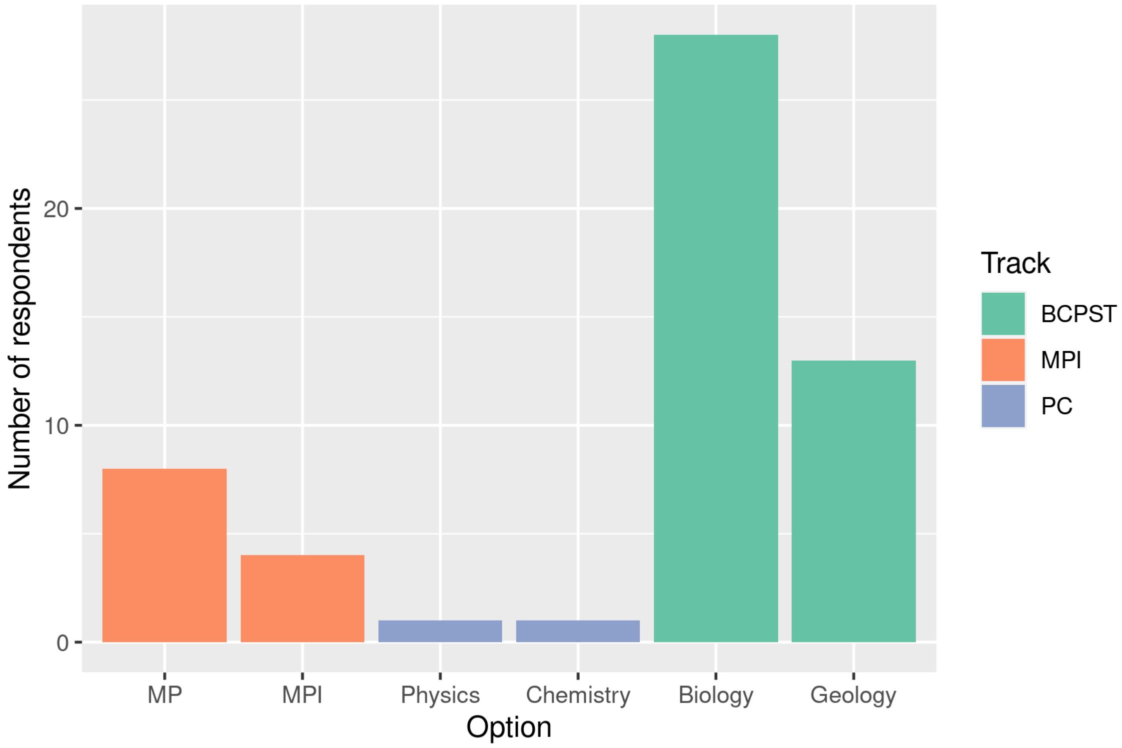


Figure 9: Distribution of the option choice of respondents, for all candidates

Option choices of the respondents The option chosen by the respondents, by track, are plotted in Figure 9. Except for the PC track, which is not representative because of the few number of answers, they are quite close to the reality of the choice of option: two-thirds of BCPST candidates chose Biology, which is a little bit under the average for eligible candidates, for the years 2016-2019 (76%). In MP track, the gap is even closer: one-third of the respondents chose MPI, while 30% of eligible candidates chose MPI between the years 2016 and 2019. For both the BCPST and the MPI track, our data seems to be representative of the population of eligible candidates.

4 Theoretical framework: a double-sided model explaining choices of students and the reaction of the school

In order to better understand the choices of candidates in this context, we made a two-sided model, that can be separated in two parts. In subsection 4.1, we wrote a short model of choice from the students' side; we also discuss what possible interpretations are behind "gendered stereotypes". The, in subsection 4.2, we model one possible reaction of the ENS when facing these gendered-determined choices from students; this part is heavily inspired by Breda and Ly, 2015, which empirically demonstrates that the school is responsible for gendered positive discrimination. Finally, in subsection 4.3, we compute both sides of the model together.[7]

[7]I would like to thank Thierry Verdier, who gave the Social Interaction seminar last year, and Yoann Perdereau, who worked with me on this model for this seminar; their input and their help were very valuable.

4.1 The "offer" side: the students' side

4.1.1 Definition of a gendered-topic, and of social sanction

Gendered-topic We want to consider an educational context where students have to choose between two topics, to which are associated a variety of stereotypes linked to gender (what we will call "gendered-topics"; in the most simple version of this model, that we will explore here, there is one feminine and one masculine topic). However, it seems important to discuss shortly what are these stereotypes, how they are conveyed to the students who have to make a choice, and what social sanctions they may imply. As we discussed in 2.2, it is rather clear that there is a gendered segregation across topics (the more mathematics-intensive ones being the more masculine ones), and that there exists beliefs that associate some genders with some fields and competences. However, if these two facts are correlated, it is not easy to understand the correlation at work: are girls less encouraged to study mathematics because of beliefs that they cannot perform as well as boys, or are beliefs the consequence of the observation that there are no girls in mathematics? Moreover, empirically, it is hard to estimate the "strength" of a stereotype within a specific population (here, the candidates to scientific entry track to the ENS, which are likely to be specific because this exam is highly selective and because the candidates are already specialized in scientific topics). One possibility to estimate the strength of the stereotype associated with a topic is to take an exogenous variable as a measure: this is what has been done by Breda and Ly, 2015, where the indicator taken was the proportion of women professors in higher education in France. In this model, we choose a different and complementary approach, by taking the share of women taking a topic in a given track as an endogenous measure of the stereotype. The interest of this measure is that the share of women in a topic is directly observable by other students, and is hence likely to be a good measure of their social context, while avoiding the difficulty of defining explicitly the stereotypes.

Social sanction After having defined a measure for the stereotypes associated with a topic, it is important to spend some time on defining and explaining the social sanction associated with these stereotypes. As we saw in the literature review, social sanctions can take two forms: external and internal. An external social sanction could, for instance, be a teacher that states regularly that women are less proficient than men in his field; the social sanction could also be the consequence of the behavior of the peers, the family, or of the ENS. Social sanctions can also be internalized; in this case, the person would not face any sanction after deviating from the social norm, but nonetheless act as if this sanction exists. One possible interpretation of this internalized sanction is that individuals may be underestimating their real level in a gendered topic that does not correspond to their gender. Of course, both forms of sanction can coexist at the same time, and reinforce each other; our model will hence have to take this generality into account when defining a social sanction.

4.1.2 Setting of the model

Specifications of the model We first assume that there are two genders, men and women, which we represent by the index $i \in \{\mathrm{M}, \mathrm{W}\}$, and that these genders are equally distributed in the student population. Each student chooses a subject between a feminine-topic (A) and a masculine-topic (B). We assume that the stereotypes, which are preexisting to the agents and imposes itself on them, creates a social norm, and that any deviation from this norm is subject to a social sanction S (that can be external or internal). This social sanction directly reduces the usefulness of the student if he or she chooses the gender deviant subject.

We assume that the student's choice of orientation is made according to a utility function, which depends on both his or her gender and the subject chosen. The student positions himself in such a way as to maximize the grade he will get in the subject, while also seeking to avoid the social sanction associated with the norm. The utility of a student of type i who chooses the topic T can then be expressed in the following form:

$$U(i,T) = \mathbb{E}[\text{Grade}_T] - S_{\text{i,T}}\mathbb{1}_{\text{choice}}$$

Thus, utility depends on two variables: the variable $\mathbb{E}[\text{Grade}_T]$, which is the student's expected grade in topic T, and the social sanction $S_{\text{i,T}}$ which takes effect with the indicator $\mathbb{1}_{\text{choice}\neq\text{norm}}$ if the subject choice does not meet the norm. For example, a woman choosing a feminine topic A receives a utility

$$U(\text{w}, \text{A}) = \mathbb{E}[\text{Note}_{\text{A}}]$$

while a woman choosing a masculine topic B receives a utility

$$U(\text{w}, \text{B}) = \mathbb{E}[\text{Grade}_{\text{B}}] - S_{\text{w,B}}$$

Thus, the comparison of these two utility functions guides the student to a subject, according to three scenarios. The first is the easier scenario: if the student is better in the subject that corresponds to his or her gender, then he or she will choose this subject regardless of the social sanction, and this choice will be efficient. The second scenario is the "efficient deviant" ("déviant efficace") scenario: if a student is much better in the deviant subject than in the normal subject, and this difference is sufficient to compensate for the social sanction, then he or she will choose the deviant subject, but this choice will be socially efficient because the student's talents will be allocated to the right place. In contrast, the distortion arises from the third "inefficient" scenario: if the student is slightly better in the deviant subject, but not enough to compensate for the social sanction, then he will choose the subject associated with his gender, and this will result in an inefficient allocation of resources. The three scenarios are illustrated by the Figure 10.

Having introduced the utility function, we now detail the expression of its two component variables: the grade and the social sanction. The grade depends on three parameters: the student's level, the discrimination applied by the examiner and a random part. The random part, denoted ε, comes from the fact that there is always a chance component in an exam, but also from the fact that the examiner imperfectly observes the student's n_T level in the T topic. The formula for the grade for a student of type i in topic T is therefore

$$\text{Grade}(i,T) = n_T + d_{\text{i,T}} + \varepsilon$$

The variable $d_{\text{i,T}}$ corresponds to the discrimination to which the student i is subjected in the topic T. This discrimination has been added to allow for grading that systematically (i.e. non-randomly) differs between students; we will endogenise it in the following subsection.

Finally, we decide to endogenize the social sanction $S_{\text{i,T}}$. Since it is the sanction for deviating from the norm, it must increase with the strength of the gender norms: the more differentiated a situation is, the higher the social sanction and the more it allows the reproduction of this differentiated norm. Thus, we would like the social sanction for girls choosing mathematics to decrease with the number of girls in mathematics, and vice versa for boys choosing literature. Moreover, we want the social sanction to cancel out if there are as many boys as girls in a subject, because in this case there is no norm and therefore no sanction. Thus, we opt for the following form: the social sanction for a girl a masculine topic B is equal to

$$S_{\text{w, B}} = \beta(r_{\text{B}} - 0.5)$$

with r_{B} the ratio men/women in topic B (the closer it is to 1, the more boys there are). Conversely,

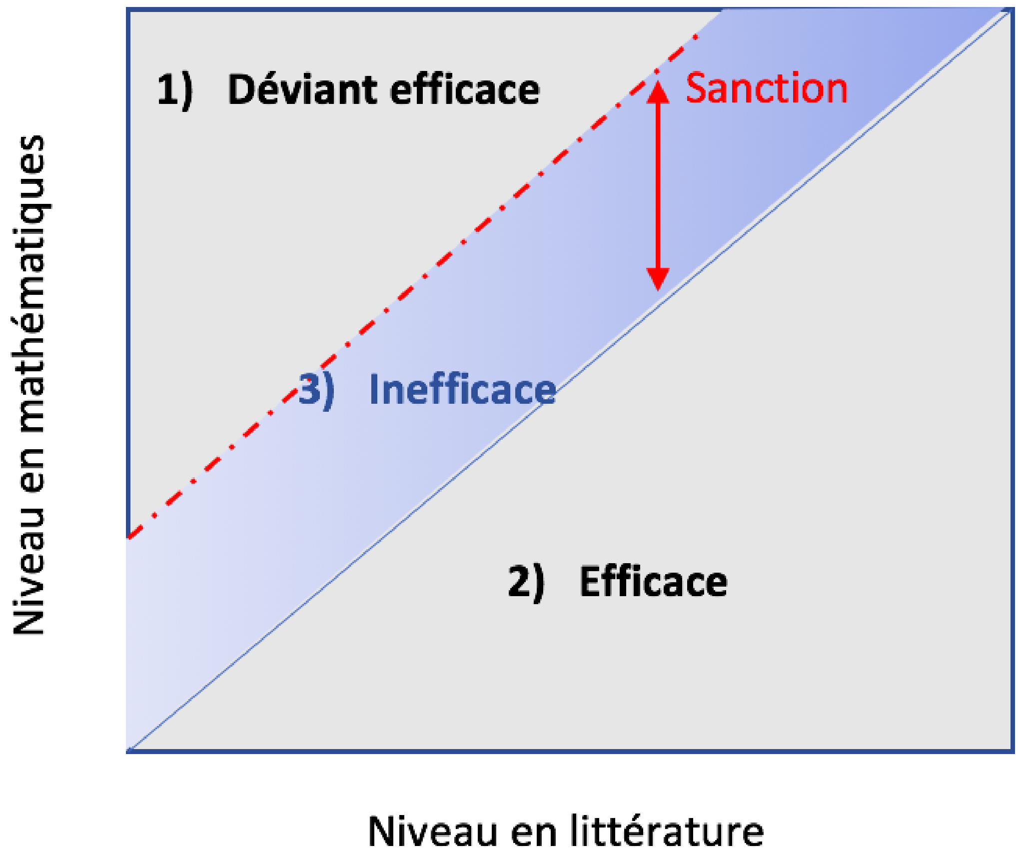


Figure 10: Representation of the three possible scenarios when deviating from a social norm

the social sanction for a boy choosing the feminine topic A is equal to $S_{\text{m, A}} = \beta(0.5 - r_{\text{A}})$, with r_{A} the men/women ratio in topic B. There are several things to note about this function. First, we consider here and in the rest of the paper that the reference ratio is the men/women ratio, not the women/men ratio. By abuse of notation, our r_{B} therefore actually means $r_{\text{m,B}}$, which is equal, under the assumption that there are as many women as men, to $1 - r_{\text{w,B}}$. Conversely, we have $r_{\text{A}} = r_{\text{m,A}} = 1 - r_{\text{w,A}}$. Then, we consider that the positive discrimination for men in feminine topics is equal to the positive discrimination for girls in masculine topics. In our symmetric men-women framework, this assumption makes sense: we consider no other societal discrimination here, and men and women are otherwise equivalent. Finally, if we logically consider that r_{B} is greater than 0.5, i.e. that there are indeed more men than women in masculine topics, this means that our function is defined from 0.5 (where it is zero, the social sanction cancels out because there is no norm) to 1, where it is at its maximum, because there are 100% boys in mathematics, and therefore the norm and the social sanction are high. The social sanction is thus a linear function, increasing and contained in the interval $\left[0, \frac{\beta}{2}\right]$.

Thus, by combining the previous specifications, the utility function for a woman choosing a feminine topic can be written

$$U(\text{w}, U(\text{w}, \text{A}) = \mathbb{E}[n_{\text{A}} + \varepsilon]$$

while for a girl choosing a masculine topic it is equal to

$$U(\text{w}, \text{B}) = \mathbb{E}[n_{\text{B}} + d_{\text{w,B}} + \varepsilon] - \beta(r_{\text{B}} - 0.5)$$

All that is left is to derive the fixed point of this model, that is, the stable fraction of men and women in each subject.

Results The fixed point in the model, if it exists, is the men/women ratio such that, for this ratio, the social sanction has a certain value and, for this value of social sanction, students make their choices in such a way as to obtain the same ratio, $r_{\mathrm{B}}\left(S_{\mathrm{w,\ A}}(r_{\mathrm{A}})\right) = r_{\mathrm{B}}$. Since our reference ratio is the men/women ratio, we will focus our analysis on the choice of men. However, we should keep in mind that there are as many men as women in this model: thus, we have $r_{\mathrm{m,B}} = 1 - r_{\mathrm{w,B}} = r_{\mathrm{w,A}} = 1 - r_{\mathrm{m,A}}$, so we only need to find one ratio, here r_{B}, to find the others (of course, we can slightly ameliorate the model to take into account different share of men and women). Using the previous formulas, we can find the choice of a student: thus, a man will choose the masculine topic B if and only if

$$U(\mathrm{m},\mathrm{B}) \geq U(\mathrm{m},\mathrm{A})$$

i.e. if and only if

$$\mathbb{E}[n_{\mathrm{B}} + \varepsilon] \geq \mathbb{E}[n_{\mathrm{A}} + d_{\mathrm{m,A}} + \varepsilon] - \beta(r_{\mathrm{B}} - 0.5)$$

Since there are as many boys as girls in our model, we obtain the following equilibrium ratio:

$$r_{\mathrm{B}} = \mathrm{Prob}\left(n_{\mathrm{B}} - n_{\mathrm{A}}\right) \geq d_{\mathrm{m,A}} - \beta(r_{\mathrm{B}} - 0.5))$$

It is possible to solve this model analytically: the computations for a uniform distribution of grades are available in appendix B.1. The computations yield quite simple results once the two parameters, $d_{\mathrm{m,B}}$ and β, are calibrated to "realistic" values since our model is very stylized. Thus, if we assume that each individual who specialize in the topic opposed to their gender receives a positive discrimination of $d_{\mathrm{m,B}} = 2$ points, and that the parameter β of the social sanction is equal to 40 (which means that, if there are 60% of boys in mathematics, then the social sanction for a girl is 4), the previous ratio is equal to

$$r_{\mathrm{B}} = 0.606$$

Thus, this calibration results in a man/woman ratio in the masculine field of 61%. For different values of social sanction and discrimination, and provided that the social sanction is always greater than or equal to the positive discrimination, the ratio evolves between 0.5 (no differentiation) and 0.9 (very great differentiation). As an indication, we show the graph of the boy/girl ratio as a function of discrimination and social sanction in Figure 11

Thus, the ratio logically increases with the gap between social sanction and discrimination. If the social sanction for a girl in mathematics is 10 points and the positive discrimination is zero, then 87% of the students in mathematics will be boys, and only the girls who have a difference in level between mathematics and literature greater than 10 points would choose mathematics. Conversely, if positive discrimination is equal to the social sanction (we do not consider the case where it is greater), the ratio will be 50% and there will be no differentiation between topics.

4.2 The "demand" side: the school's side

In the previous subsection, we introduced the variable $d_{\mathrm{i,T}}$, which corresponds to the discrimination to which the student i is subjected in the subject T. This discrimination can be conscious or not on the part of the examiner, but is not necessarily negative: it can also be "positive discrimination", i.e. an addition of points to compensate for the social sanction. Here, without loss of generality, we consider only positive discrimination, and assume that either teachers do not apply negative points according to gender, or that negative discrimination is already included in the social sanction. Thus, we have $d_{\mathrm{w,A}} = d_{\mathrm{m,B}} = 0$

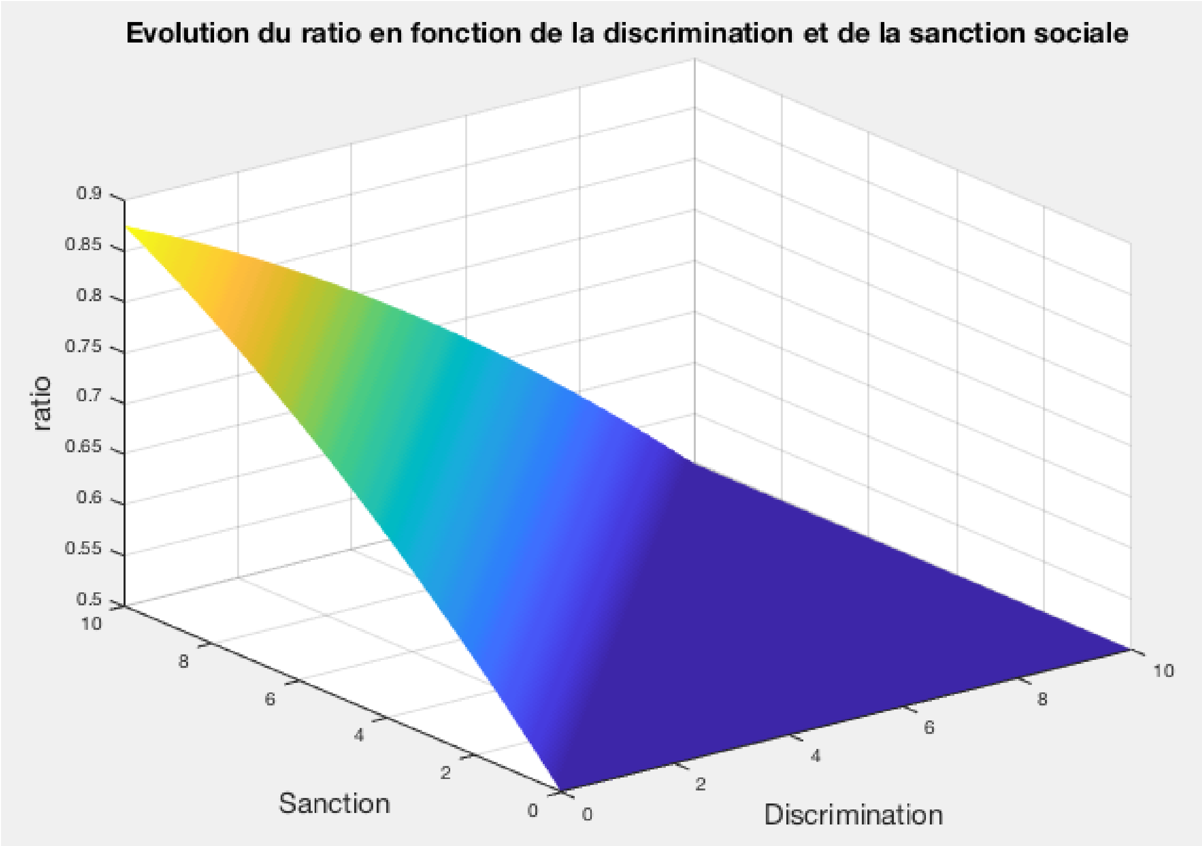


Figure 11: Evolution of the men/women ration in regard to the social sanction and to the positive discrimination

(no negative sanction for students who choose the subject associated with their gender), and we have $d_{\text{w,B}}, d_{\text{m,A}} \geq 0$ (positive discrimination towards students who choose the deviant subject). From the student's point of view, the positive discrimination was exogenous: however, this subsection is dedicated to the endogenisation of this social sanction, by providing a rational explanation of why the school may want to apply such positive discrimination.

The school aims to recruit the best students, who are all evaluated on two subjects, A and B. However, the school does not wish to favor a solely generalist profile; on the contrary, it wishes to recruit people who are at least partially specialized in one of the two proposed subjects. To do this, the school calculates the candidates' grade (which will then determine their admission) by weighting the subject in which the student is the best with a higher weight: this coefficient, α, is therefore greater than 1. The student's level, n, (for the school) is equal to :

$$n = \max(n_A + \alpha n_B; \alpha n_A + n_B)$$

For example, for a student who is better at B than A, i.e., $n_B > n_A$, the student's level for the school is:

$$n = n_A + \alpha n_B$$

However, the choice of option is left to the students, who can choose either A or B as their major. Once the student has chosen, the school assigns the grades (N_A, N_B) that it wants based on the level (n_A, n_B) it observes in subjects A and B. It is assumed that the level is observed perfectly by the school. It is also assumed that each student is graded independently of the others, and that their A and B grades are themselves independent. The overall grade assigned by the school, on which the students will be ranked, is therefore, depending on whether the student chose option A or B :

$$N = N_A + \alpha N_B \text{ or } N = \alpha N_A + N_B$$

The school's objective is therefore to minimize a cost function, which corresponds to the difference between the student's level and the final grade obtained by this student, for each student i[8]. :

$$C_i = (N - n)^2$$

The school's program for the student is therefore:

$$\min_{N_A, N_B} C_i \text{ knowing } n_A, n_B, \alpha$$

No social norm In the absence of a social norm, the school takes into account the fact that students are purely rational agents, who make their decisions by seeking to maximize the grade they will obtain: it therefore knows with certainty that the student chooses to put a greater weight on the subject on which he is the best. This is true only under the assumption that the choice of option has no consequence on anything except for the exam, which is the case: any student entered with any option can choose any department once having entered the ENS. In this case, the minimization of the school's cost function has an obvious solution: $N_A = n_A$ and $N_B = n_B$, regardless of the student's talent or gender.

Presence of a social norm In the case where a social norm exists, the behavior of students changes: they will no longer necessarily choose as an option the subject in which they have the highest proficiency. Three cases are then possible for the school, depending on the level of student i, who is socially penalized when he chooses option A:

- $n_B > n_A$, or "logical" scenario: this is the case where the student is better in the subject that corresponds to her social norm (for example, a girl who is better in subjects perceived as feminine). In this case, the student will obviously choose B as an option, and the school has no discrimination to implement.
- $n_A - S_i > n_B$, or "effective deviant" scenario: in this case, the student is better in the subject that does not correspond to his gender, and this difference in level more than compensates for the cost of the social sanction. Student i then chooses A as an option, which is the subject in which he is the best: the school thus has no discrimination to implement.
- $n_A - S_i < n_B < n_A$, or "inefficient" scenario: in this case, the student is better in A, so the school would like him to choose subject A as an option; but as this student is socially penalized when he chooses subject A, and the weight of this penalty is too high, he will deviate from the optimal behavior from the school's point of view, and will then choose B as an option, despite his lower level in B.

It is in this last case that there is a discrepancy between the student's grade and his level according to the school. The student's level according to the school is :

$$n = n_B + \alpha n_A$$

[8]As all grades are independent, the school minimizes each cost function independently; we can therefore be satisfied with a single student

But the grade that will be given will be:

$$N = \alpha N_B + N_A$$

It is thus easy to see that giving a grade equal to the level is no longer an optimal strategy for the school, which will want to correct the students' bias by giving a different grade. To obtain the school's strategy, we simply minimize the cost function established above. We assume, as in part 1, that the school only discriminates in the subject in which the social norm is operative (here, this means that for individual i, $d_{i,B} = 0$). We then find:

$$\begin{aligned} \alpha N_B &= n_B + \alpha n_A - N_A \\ \iff N_A &= n_A + n_A(\alpha - 1) + n_B(1 - \alpha) \\ \implies d_{i,A} &= n_A(\alpha - 1) + n_B(1 - \alpha) \\ &= (n_A - n_B)(\alpha - 1) > 0 \end{aligned}$$

Thus, we see that the school will have to impose positive discrimination on certain students if it wants to correct the impact that the existence of a social norm has on its recruitment. This positive discrimination is offered to students who are led to modify their choice of options when a social norm is introduced, and it is all the more important for a given student when the gap in levels between the two subjects is large.

4.3 Unification of the two sides of the model

In the first subsection(4.1), we modeled students' choices of subject according to a social sanction imposed on them and considering an exogenous positive discrimination. The social sanction was endogenous, and depended directly on the men/women ratio in each subject. Then, in a second subsection(4.2), we modeled the choice of schools, which try to attract the best candidates but know that students may be forced into inefficient choices because of the social sanction. This distortion leads schools to positively discriminate for students in a gender deviant subject. In this third part, we will now combine the two sides of this model, to find the fixed point.

Reuniting the two sides of the model has no easy analytical solution. Thus, we created an algorithm that simulates a large number of students, and randomly assigns them a gender and a level. To make the simulation more realistic, the level is assigned following a normal distribution $N(10, 5)$, which was not possible in its analytical version (in appendix B.1). The algorithm, for a certain value of the coefficient α, starts from a men/women ratio in each given subject, calculates the positive discrimination and the social sanction implied by this ratio, then simulates the choice of the students, which creates a new men/women ratio in each subject. The algorithm is repeated as long as the new ratio is different from the old one, and converges to a fixed point. Moreover, we compute this algorithm for different values of α, the coefficient set by the school. Thus, we obtain the boy/girl ratio in mathematics for different values of α, which we reproduce in the Figure 12. Note that the men/women ratio in masculine field is equal to the women/men ratio in feminine field, because there are as many women as men in our simulation and because we conceive our model as symmetrical.

As we can see, the equilibrium men/women ratio decreases with the coefficient α. Indeed, as the optimal positive discrimination of the school is $d_{i,A} = (n_A - n_B)(\alpha - 1)$, it increases with α. Thus, if the school puts a high coefficient on a subject, the differences in level in this subject result in larger and larger differences in the grade, which reduces the inefficiency zone. Thus, increasing the coefficient increases the magnitude of the grade difference, which is therefore more likely to be greater than the social penalty,

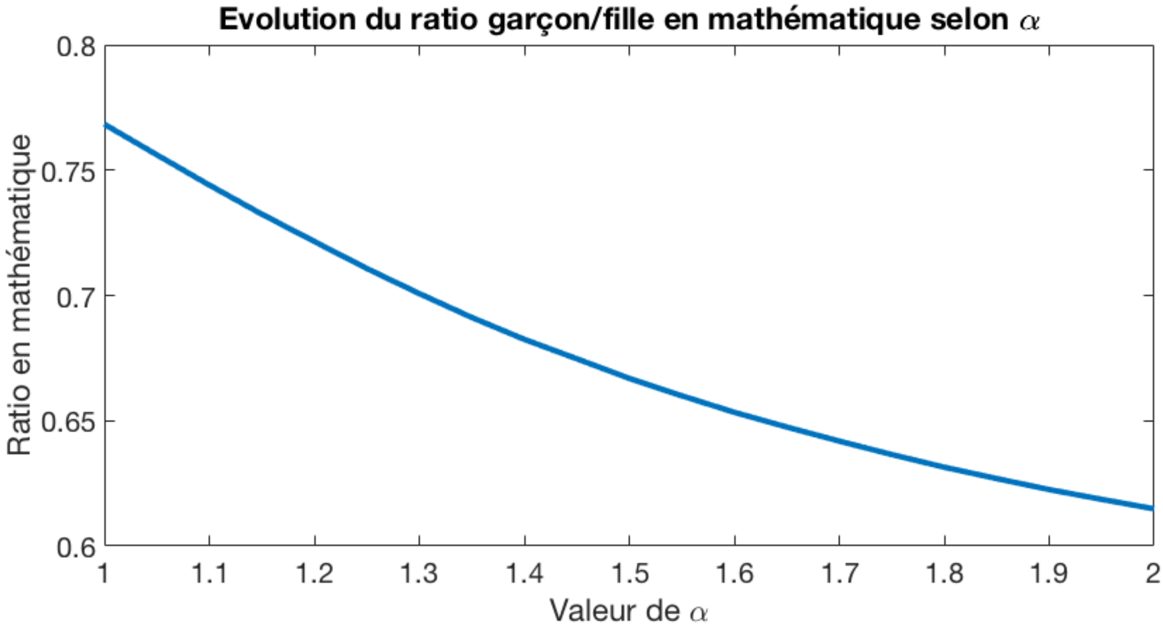


Figure 12: Evolution of the men/women ratio (and hence of the social sanction) according to the relative weight the school puts on specialization topic

thus reducing the bias. For example, a coefficient $\alpha = 10$ implies a men/women ratio in mathematics of 52%. In extreme cases, a ratio that would be infinitely large would completely reduce the distortion. However, it should be remembered that the coefficient is not only there to create a positive distortion, but above all to select students who have a strong point. A very large α coefficient would mean focusing on only one subject, which is also not what the school wants, as it is looking for a polyvalent profile, but with a strong point. Thus, the α coefficient decreases with the desire to have polyvalent profiles, and increases with the desire to look for students with strengths and to reduce the bias.

The choice to increase the diff between two options has been opted by the ENS for the PC track: beginning in the 2022 session of the exam, students who specialize in physics also have higher weights on the mathematics exam (both written and oral), while the chemistry track put less weight on physics and mathematics and more on chemistry[9]. According to our model, these changes of weights may have an impact on the gender-segregation in option choices; an extension of this paper would have been to test this hypothesis with the 2022 data of exams, but these data were not accessible before the writing of this paper.

5 Methodology

5.1 Identification strategy

As explained in section 2.1, the ENS entrance exams offer a very particular setting for two of the entrance tracks: PC and BCPST. For these two tracks, students have to choose an option (between Physics and Chemistry for the PC track, and between Biology and Geology for the BCPST one). This choice of option should be based on a purely rational decision: in fact, these options does not impact the curriculum that are followed in CPGE, nor do they impact the studies at the ENS if the candidate is admitted. In addition to that, the goal of CPGE is to enter the best school possible; it is then very likely that candidates are informed on this choice, and that they are taking the choice of option as an important decision. As a

[9]The new weights are reported at https://www.ens.psl.eu/sites/default/files/arrete_7-12-2021_boesr_23-12-2021.pdf

consequence, in this setting, one could expect that the gender of the candidates should not impact in any way the choice of option, if this choice is, as predicted, rational. The goal of our identification strategy is to test for this hypothesis, by doing a probit regression of the likelihood of choosing an option on the level of the candidate in both topics. This is made possible by our context: we have a simultaneous measure of the level of the candidates in the two topics that can be chosen as an option. Our regression equation can hence be written as :

$$P_i(Y=1) = \phi(\beta_0 + \beta_1 \mathbb{1}_{\text{i = women}} + \beta_2 G_{i,Y=1} + \beta_3 G_{i,Y=0})$$

Where Y is a dummy that is equal to 1 where the topic considered is the "masculine" topic, and 0 where the topic considered is the "feminine" topic; $G_{i,Y}$ stands for the grade that the candidate i gets in topic Y; and $\mathbb{1}_{\text{i = women}}$ is a dummy variable that is equal to 1 when i is a woman. The coefficient of interest is β_1; under the hypothesis that candidates choose rationally their options, this coefficient should be 0. However, if we can statistically reject the null hypothesis for β_1, that means that the influence of gender on the option choice is significant, and hence that the choice of option is not purely rational. The results of this specification are visible in Table 5 and Table 6.

5.2 Counterfactuals

A crucial feature of our empirical setting is that all candidates are taking all the exams, regardless of their option. This allows us to compute what we call "counterfactual results", i.e. the result one candidate would have gotten if he or she had taken the other option. In order to do this computation, we simply recompute the final grade that one candidate gets by changing the weights that are put on each exam, which are reported in Table 2. It is then possible to generate the maximum result one student could have obtained, which is equal to her actual result if she chose the option where she was stronger (the "right" option), and is equal to the counterfactual result otherwise. This maximum results can then be used to study the gender composition of the admitted candidates; in particular, it allows us to study whether the gender composition of the admitted candidates, and their level, is dependent on individuals choosing the right option, and what would happen if no option existed. These results are presented in section 6.2.

5.3 Impact on the choice of department for admitted candidates

Our final goal is to understand whether the choice of option during the exam has an impact on the choice of department once in the ENS, for the admitted students. Once again, since admitted students are free to choose any department they like, one would predict that their department choice should not be influenced by their option choice, which is only an exam artifact. In order to test this prediction, we run a probit model similar to the one described above; however, the dependent variable is now the entrance department of the student i (precisely, it's a dummy variable that equals one when the department chosen is the "masculine one" (physics for the PC, geology for the BCPST), and 0 if it's the feminine one; we are excluding from our dataset students who go in a different department, because they are very few of them, as shown in Table 4).

Our identification strategy is based on the hypothesis that students know that their option does not influence the department they will study in after enrollment; if this was not true, our regression could no longer be interpreted as causal, because students may have chosen their options according to what department they would like to study in (reverse causality). We will discuss the validity of this assumption in section 6.4.

6 Results

6.1 Choice of option with regards with gender, by track

First, we estimate the impact of the gender of the candidate on the option chosen at the written stage of the exam (for all candidates) and at the oral stage (eligible candidates only); we then compare and interpret these two estimations.

6.1.1 Written grades

The estimation of the specification on all the candidates at the written exam stage, presented in section 5.1, is displayed in Table 5, for the PC track (columns (1) and (2)) and the BCPST track (columns (3) and (4)).

Table 5: Prediction of option choice according to the gender and the grades, for the written exams (all candidates)

	Dependent variable:			
	Likelihood of choosing physics		Likelihood of choosing geology	
	(1)	(2)	(3)	(4)
Woman	-0.575^{***}	-0.225^{***}	-0.230^{**}	-0.017^{***}
	(0.083)	(0.032)	(0.092)	(0.006)
Grade in Chemistry	-0.185^{***}	-0.071^{***}		
	(0.015)	(0.006)		
Grade in Physics	0.133^{***}	0.051^{***}		
	(0.012)	(0.005)		
Grade in biology			0.029^{*}	0.002^{*}
			(0.015)	(0.001)
Grade in geology			0.009	0.001
			(0.015)	(0.001)
Constant	1.029^{***}		1.691^{***}	
	(0.202)		(0.100)	
Observations	1,470	1,470	3,037	3,037
Log Likelihood	-852.350	-852.350	-463.315	-463.315
Akaike Inf. Crit.	1,712.699	1,712.699	934.630	934.630

*p<0.1; **p<0.05; ***p<0.01

Note: columns (1) and (3) are the results of the probit regressions for the PC and the BCPST track respectively; columns (2) and (4) are the marginal effects of regression (1) and (3), estimated at average

As we can see, the gender has a significant impact on the likelihood of choosing one option rather than the other, and this is true in both tracks, although the effect is much larger in PC than in BCPST. Moreover, in both cases, this effect is in the same direction as the gender norms; women are less likely to choose physics and geology, and men are less likely to choose chemistry and biology. Moreover, the size of this effect is very large for students in the PC track: on average, a woman candidate has a lower probability to choose physics by 22%, which corresponds to the magnitude of scoring better at physics by 4 points, or worse at chemistry by 3 points (out of 20). Comparatively, the probability (estimated at average) for a woman to choose geology in BCPST is lower than for men, but the size of the effect if much smaller (below 2%). These results are replicated in the appendix, by replacing the grades with the rank of the students (in order to eliminate any effect that may be driven by grades differences between different years, since the rank is computed for each year independently); they are almost identical (see Table 10).

6.1.2 Oral grades

The estimation of the specification on all the candidates at the written exam stage, presented in section 5.1, is displayed in Table 5, for the PC track (columns (1) and (2)) and the BCPST track (columns (3) and (4)).

Table 6: Prediction of option choice according to the gender and the grades, for the oral exam (eligible candidates only)

	Dependent variable:			
	Likelihood of choosing physics		Likelihood of choosing geology	
	(1)	(2)	(3)	(4)
Woman	−0.397**	−0.154**	−0.095	−0.022
	(0.172)	(0.068)	(0.190)	(0.044)
Grade in Chemistry	−0.053***	−0.020***		
	(0.019)	(0.007)		
Grade in Physics	0.058***	0.022***		
	(0.018)	(0.007)		
Grade in biology			−0.038	−0.009
			(0.026)	(0.006)
Grade in geology			0.081***	0.018***
			(0.026)	(0.006)
Constant	0.357		−1.517***	
	(0.286)		(0.411)	
Observations	402	402	268	268
Log Likelihood	-251.931	-251.931	-110.741	-110.741
Akaike Inf. Crit.	511.862	511.862	229.483	229.483

*p<0.1; **p<0.05; ***p<0.01

Note: columns (1) and (3) are the results of the probit regressions for the PC and the BCPST track respectively; columns (2) and (4) are the marginal effects of regression (1) and (3), estimated at average

This time, we see that while the result is still significant for the PC track, it is not longer statistically different from 0 in the BCPST track. One possible explanation for this across-track difference is that the PC track is much more segregated in terms of gender than the BCPST track; in fact, women make up for 60% of the BCPST students, while men account for 76% of PC students (see Table 1). This may suggest that the gender stereotypes are stronger in the more segregated tracks, and hence that students are more exposed to gendered social norms in tracks where there is a strong imbalance between the shares of men and women. Once again, these results are replicated in the appendix, by replacing the grades with the rank of the students; they are almost identical (see Table 11).

6.1.3 Comparison between written and oral exams, and the role of positive discrimination

Interestingly, the effect of social norms on the choice of options seems to be smaller for oral exams as compared to written exams. This difference can be explained by two hypotheses. First, the composition of the sample is very different in both cases: in fact, anyone can take the written exams, while the oral exams are taken only by the candidates who did the best scores at written exams (for reference, eligible candidates represent less than 10% of all candidates in both PC and BCPST tracks). A first possibility is that while the candidates are influenced by social norms when taking their options, the candidates who reach the oral selection tend to be better at choosing their options (that can goes either way: either the best candidates are less influenced by social norms, or the written exam is so competitive that choosing the "wrong" option does not allow candidates to access the oral exams). In order to test for this hypothesis, we can do the same regression done in Table 5 (prediction of the likelihood of choosing an option based on the written grades of the candidates), but only for eligible candidates. The results of this regression are in appendix, in Table 12. This table shows nuanced results for the PC and the BCPST tracks. In fact, we can see that the effect of gender on the choice of option is no longer significantly different from 0 for the BCPST track; this means that the wrong choice of option was entirely driven by non-eligible candidates, and that the best candidates don't seem to be affected by social norms when choosing their options. On the opposite, for the PC track, the results are identical for all candidates and for eligible candidates; the best candidates hence seem to be as influenced by social norms when choosing their option as others.

For the PC track, another explanation for the reduction of the impact of gender on the choice of option between the written and the oral exams is the positive discrimination that was established by Breda and Ly, 2015. In fact, in our theoretical model, it is rational for the school to react to non-rational choice of options by students by applying positive discrimination. However, such discrimination is possible only for oral exams, and not for written exams, since they are anonymous. As a consequence, the observed reduction of the gendered gap (from 22% to 15%) in the choice of options may be the indication that we do not measure the level of candidates, but their level plus some positive discrimination, which (at least partly) compensate for the wrong choices of options.

6.2 Impact of the gendered choice of option on composition of admitted candidates

After having established that the option choices are impacted by the gender of the candidates, the next question is to evaluate the aggregate effect of gender norms. In fact, if the men and the women are equally influenced by gender norm, we may not see any effects of the gender norms on the aggregate proportion of women recruited[10]. To do so, we have computed counterfactuals, by recreating the final grade one student would have obtained if he or she had chosen the other option. We will firstly plot the distribution of the gaps between reel grades and counterfactuals grade, before studying the effect that using counterfactuals or suppressing all options would have on the ENS enrollment.

6.2.1 Distribution of counterfactuals

In Figure 13, we have plotted the distribution of the difference between the rank obtained by students and the rank they would have obtained if they had chosen the other option, by gender and by option.

[10]We focus on the aggregate proportion of women recruited because since we are studying STEM majors, one main policy goal is to recruit more women; however, the men proportion is easily deducible from the women's proportion.

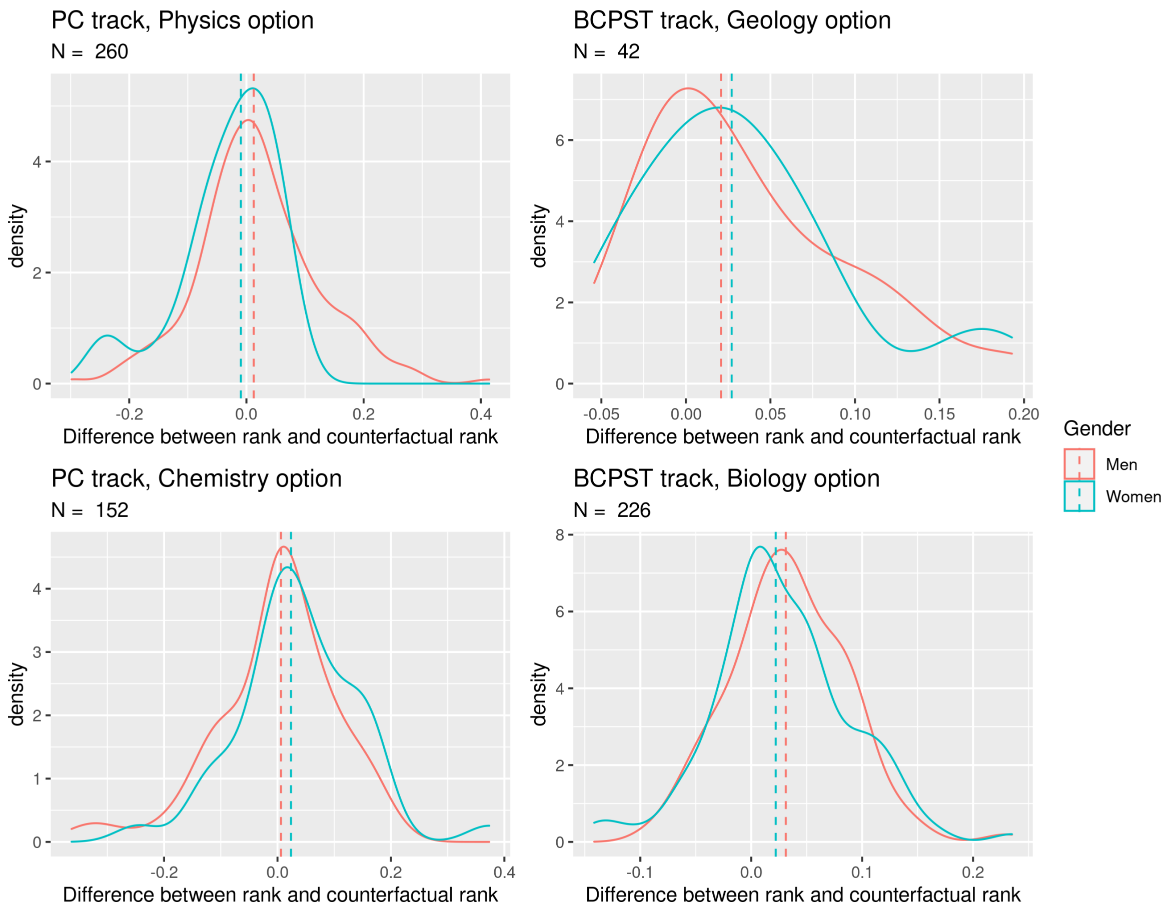


Figure 13: Distribution of the gap between the rank obtained by the students and the rank they will have if they chose the other option, for the PC and BCPST track, by gender

A negative difference means that one student was "wrong" when choosing his option, since his counterfactual grade was better than the real one; a positive difference means that the candidate chose the right option. The dotted lines represent the median of the distribution, by gender. We chose to plot the difference in the rank, rather than the difference in grade, in order to be able to compare easily across different years.

These plots reveal what seems to be different patterns for the BCPST and the PC track. For the PC track, it is interesting to note that all the distributions are centered on zero, which means that choosing the two options lead to the same final result (same grades in both topics); moreover, the distribution for the physics and the chemistry option are very similar, and it is the same for the distribution plotted by gender. At first sight, this seems at odds with the results obtained in Table 6; in fact, since there is a strong effect of gender on the choice of option, one would expect women to choose chemistry "too much" (hence the distribution for women should be on the left of the men distribution in the graph in the bottom left corner), and men to choose physics "too much" (hence the distribution for men should be on the left of the women distribution in the graph the top left corner). However, this is easily explained by the fact that men have much better grades in physics, and women better grade in chemistry (see Figure 2); the result in Table 6 has been obtained while controlling for the level, which is not the case here.

For the BCPST track, the interpretation is a bit easier. First, we note that all the medians of the distribution are positive, which means that the majority of the candidates choose the right option, which was not the case for the PC track. Second, we see that this time, the variation between the men's and the women's distributions is in line with the gendered choices of option: for the geology option, the distribution of women who choose geology is on the right than the distribution of men's in the top right corner (meaning that women need to be stronger in geology, relatively to biology, before choosing geology as an option, when compared to men), and it is the opposite for biology.

6.2.2 Impact of options on the gender of admitted candidates

After having computed the rank obtained and the counterfactual rank for each eligible candidate, we now decide to study the aggregate effect of the choice of option on the enrollment of the ENS. To do so, we compute the proportion of women in the first quartile and the first half of the distribution, in three distinct scenarios: the first one correspond to the one that really happened, the second one to the one where everyone is graded according to the option they didn't choose, and the last one where the candidates are graded according to the option where they best performed. This last scenario could correspond either to a suppression of options, or to the school deciding to put automatically a higher weight on the exam students succeeded the best. The results of these scenarios are presented in Table 7.

Table 7: Percentage of women in the top 25 and 50% of the oral exams - for the real and the counterfactuals situations

	PC	BCPST
Top 25	19%	51%
Top 25 - counterfactual	18%	52%
Top 25 - max	18%	52%
Top 50	20%	54%
Top 50 - counterfactual	21%	52%
Top 50 - max	20%	53%

As we can see in 7, the proportion of women in the first quartile and in the first half of the distribution almost doesn't change between if one consider the counterfactual scenario, and the scenario where the option is the one where the candidate has the best score. This result was logical for the BCPST track (since there was no effect of the gender on option choice at the oral stage), but is more surprising for the PC track, where gender had a strong effect. A null aggregate effect can be interpreted by the fact that the effect of gender norms on men and women compensate each other.

6.2.3 Impact of option on the level of admitted candidates

Despite having no aggregate effect on the gender of admitted candidates, having options wrongly chosen by candidates also imply that the level of the candidates admitted (as measured by the grades they obtained) is lower than the level they would have had if a higher weight was put on the best grade they got ("no option" simulation). Hence, we choose to present the distribution of the final grades in Table 8, with a renormalization of the final grades between 0 and 20.

Table 8: Summary statistics on the distribution of weighted grades, with and without options

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
PC: grade	2.801	11.289	12.393	12.258	13.560	17.769
PC: grade if no options	7.822	11.626	12.550	12.612	13.680	17.769
BCPST: grade	6.923	10.892	12.056	12.115	13.395	16.315
BCPST: grade if no options	6.923	10.945	12.128	12.201	13.534	16.562

As expected, the distribution of the grades when the higher grade is automatically considered by the school as the option is higher than the actual distribution. However, the difference is quite small for both tracks (even though it is smaller for the BCPST track), and does not seem to support the argument that the presence of option has a strong impact on the level of admitted candidates.

6.3 Long-term impact of the choice of option on the choice of major

Finally, the choice of option may also impact the choice of department of students who were finally admitted in the ENS. Two directions of causality are possible: firstly, students may take the exam while wanting to go in a specific department of the ENS (say, physics), and choose the option that correspond to that department (reverse causality). However, under the assumption that students know that their option do not condition their further choice of department, they should rationally choose their option only in order to maximize their admission probability, and this reverse causality should then not exist. We will discuss the validity of this assumption in the next question, by using our survey.

Secondly, once admitted, students who hesitate between departments may suffer from an anchoring bias, and fallback on the option they took during the exam. To test for this hypothesis, we use a similar specification as in 5.1, but with the choice of department as the dependent variable. The results are presented in Table 9.

Table 9: Prediction of department choice

	Dependent variable: department of entry							
	Likelihood of choosing physics as a major				Likelihood of choosing geology as a major			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Woman	0.341	0.100	0.482	0.155	−0.949*	−0.132*	−0.407	−0.083
	(0.586)	(0.158)	(0.515)	(0.148)	(0.532)	(0.077)	(0.377)	(0.078)
Grade in Chemistry	−0.099	−0.031	−0.141**	−0.049**				
	(0.080)	(0.025)	(0.071)	(0.025)				
Grade in Physics	0.063	0.020	0.094	0.033				
	(0.079)	(0.025)	(0.070)	(0.025)				
Option physics	1.757***	0.546***						
	(0.458)	(0.118)						
Grade in Biology					−0.113	−0.014	−0.111*	−0.022*
					(0.076)	(0.010)	(0.059)	(0.012)
Grade in Geology					−0.015	−0.002	0.067	0.013
					(0.081)	(0.010)	(0.060)	(0.012)
Option geology					2.134***	0.564***		
					(0.547)	(0.155)		
Constant	0.104		1.023		0.478		−0.326	
	(1.479)		(1.371)		(1.686)		(1.268)	
Observations	52	52	52	52	80	80	80	80
Log Likelihood	-21.117	-21.117	-29.772	-29.772	-19.245	-19.245	-29.040	-29.040
Akaike Inf. Crit.	52.233	52.233	67.543	67.543	48.490	48.490	66.080	66.080

*p<0.1; **p<0.05; ***p<0.01

Note: odd columns are the results of the probit regressions, while even columns are the marginal effects of probit regressions, estimated at average.

In Table 9, we are restricting our sample to the ENS students who choose a department corresponding to their entry track; these students represent the vast majority of each track (as shown in Table 4). Interestingly, we see that while the gender and the level of students seem to be uncorrelated with the choice of department once in the ENS, the option chosen is highly correlated with the choice of department. The dependent variable is the department of entry (department chosen by the students during their first year in the school); the same table is reproduced with the exit department, in appendix C.3, and the results are quite identical. This table suggests two possibilities: either the students choose their favorite topic as an option (for instance, because they don't know that they are free to study in every department once in the ENS - we will discuss this possibility in next section), or there is in fact an anchor bias: students may maximize their probability of entering the ENS when choosing their option, but after entering, since they have to quickly choose their department, they may choose the same department as their option. If the latter is correct, this provides an argument for the suppression of options in the entrance exams, since these options have the unwanted effect of affecting the choice of major of students.

6.4 Discussion of our hypotheses with the survey

To conclude this paper, we will examine our hypothesis in light of the survey that was distributed to this year's candidates. As a reminder, the detail of the questions and of the answers are available (in French) in appendix A.3. We will here select the core questions from our appendix in order to test for our three main hypotheses: (1) students are aware of gender norms associated to their options; (2) students are choosing their options in order to maximize their chances of passing the entrance exams and (3) students are aware that their option's choice will not affect their major choice once in the school.

6.4.1 Perception of gender norms by students

Our first working hypothesis is that students are aware of gender norms, and that the gender norms they perceive are in line with the societal gender norms associated with topics. The question 22 asked students whether the option in which they are enrolled is more associated with "masculine" or "feminine" stereotypes. The results of this question are displayed below, in Figure 14.

As we see in Figure 14, candidates are aware of gender stereotypes associated with an option: all options tend to be designated as "masculine" or "feminine" by candidates, even though a "neutral" option was offered. Additionally, candidates all agree on the stereotype associated with a topic, suggesting that these stereotypes are widely shared and not debated. Moreover, stereotypes perceived by candidates are in line with the share of women in each option (there are more women enrolled in biology as compared to geology, and biology is considered as a feminine track, while geology is considered a masculine one; the same logic applies for MP versus MPI track). This validates our first hypothesis: candidates are aware of gender norms and therefore, likely to be influenced by these norms.

6.4.2 Factors of the option choice

Our second working hypothesis is that students choose their option in order to maximize their probability of passing the exams, as a rational agent will do. When asked directly the question (question 16), most students (70%) answer that they did choose their option in order to maximize their chances of entering the ENS. This number is quite high; however, one could expect it to be even higher, considering that options have no other role than helping student to enter the ENS. When asked why they choose their option (question 15), answers are more nuanced, as we can see in Figure 15.

While students could select more than one answer to question 15, it is worthy to note that the first

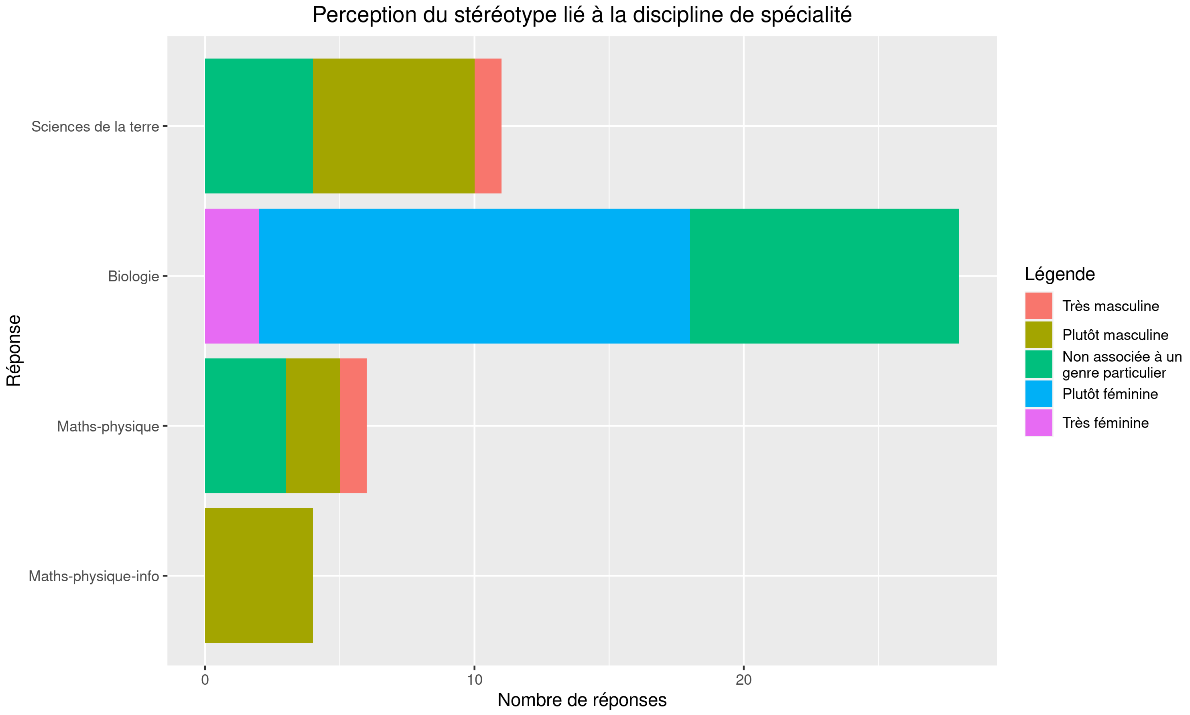


Figure 14: Perception by the candidates of the stereotypes associated with their option

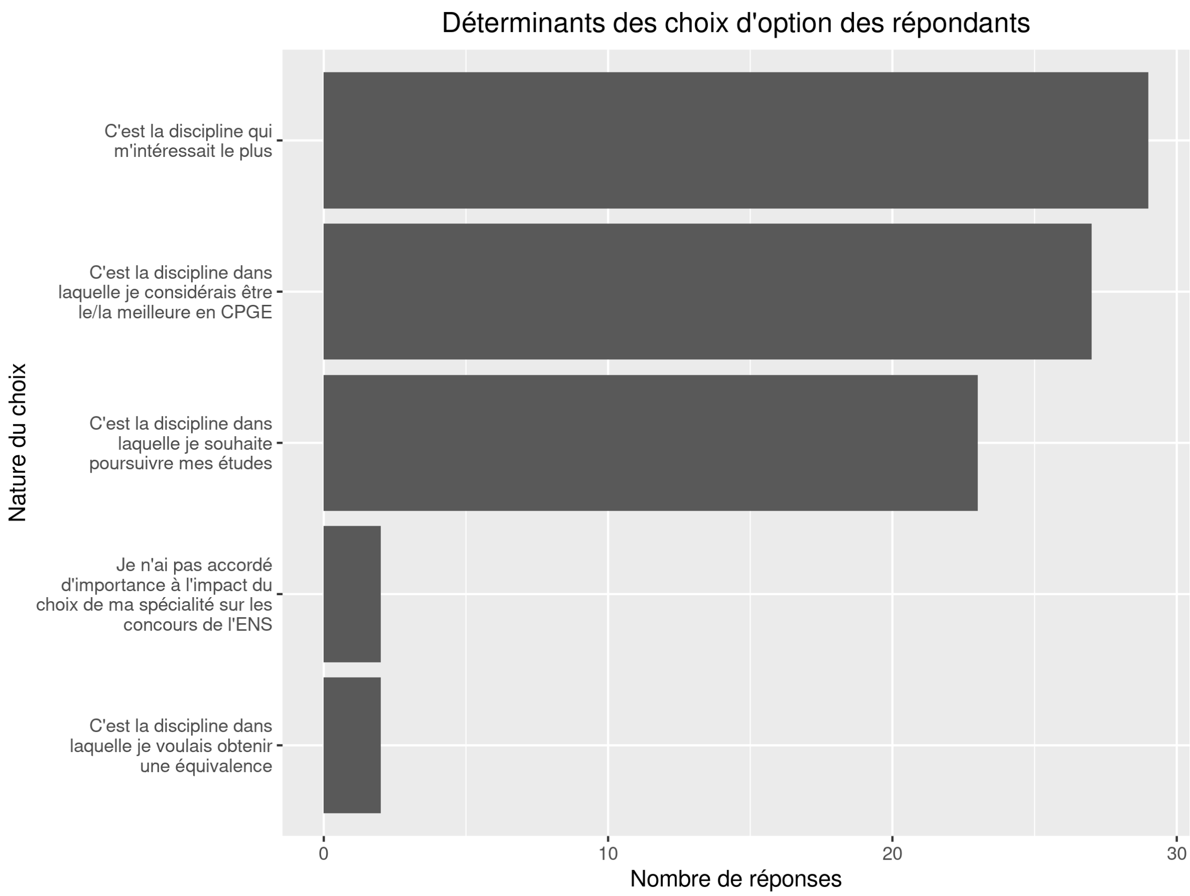


Figure 15: Factors of the option choice

factor of option choice is the interest that the students have in the discipline, with the "rational choice (choosing the topic one consider to be the best in) arrives only in second position. This suggests that despite being in an environment that incite students to be rational (they spend two years to study for an exam, they know how the exam and the options work, the goal is binary and simple), students still prefer to "go with their guts" than to be strictly rational. It would have been interesting to see if students that are explicitly making a rational choice get better results than students who do not; however, as explained before, the results of the 2022 exam session were not available in time. This is a limit of our analysis, and leave room for an improvement of this research.

6.4.3 Option choice and department choice

Finally, our last hypothesis is that students are aware that their option choice does not impact their department choice, if they enroll in the ENS. Question 23 of the survey ask the students if, according to them, their option choice will impact their department choice; their answers are displayed in Figure 16.

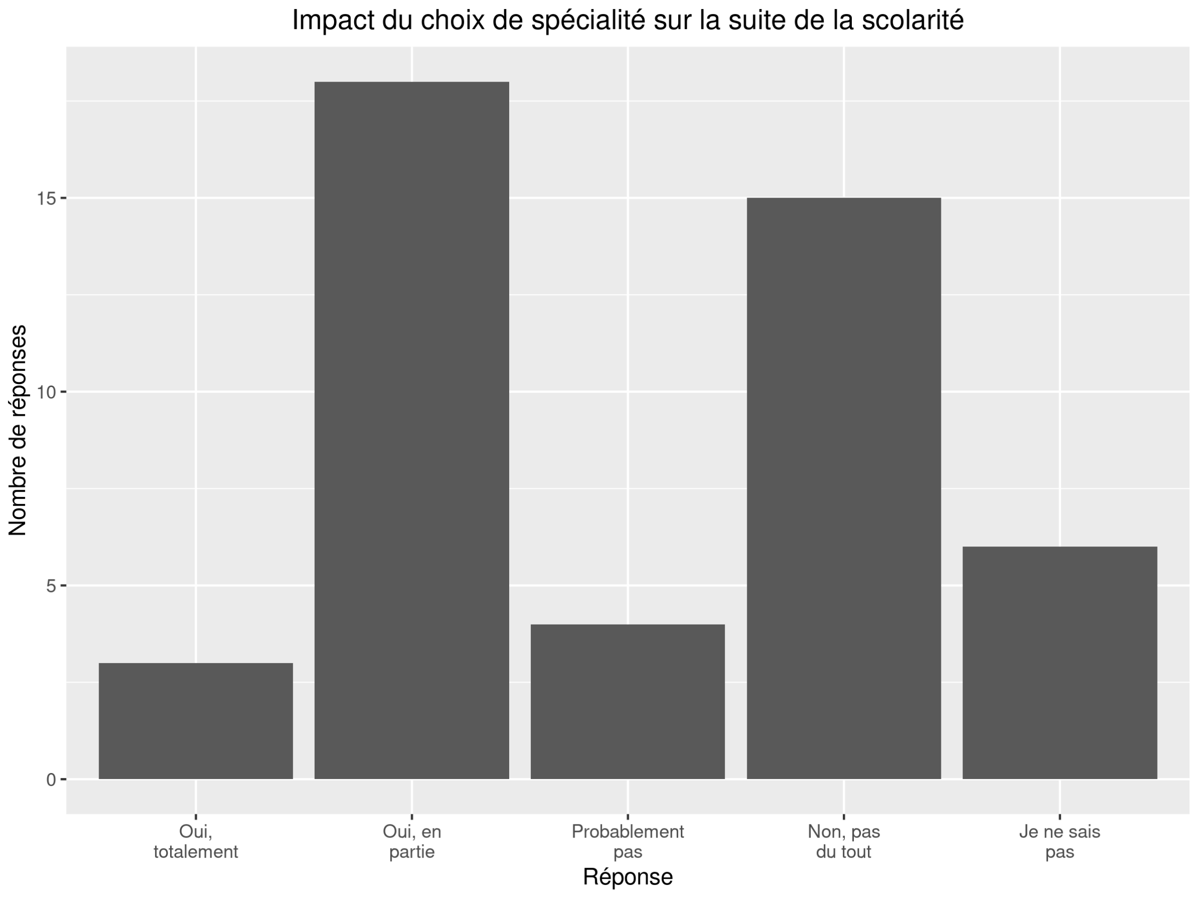


Figure 16: Option choice and department choice

Interestingly, a majority of students believes that their option choice will impact, at least partly, their academic direction. This is likely to impact our analysis; in fact, if students chose their option according to the topic they would like to major in, it is then rational for them to make their option choice according to their preferences (in addition to their abilities). This result suggests that despite the important communication of the school, more should be done in order to fully inform candidates of the modalities of the exams.

7 Conclusion

In this paper, we contributed to a measure of the role of gender norms (and, more largely, social norms), even in contexts where these norms would be likely to be weaker than in general society. We showed that the choice of option (between a more masculine stereotyped topic and a more feminine one) is influenced by gender after controlling for the level of candidates in each topic, despite the absence of long-term impact of this choice. Moreover, the impact of gender is greater in the more gender-segregated track (Physics-Chemistry, 24% of candidates are women) as compared to the less segregated track (Biology and Geology, 59% of candidates are women), which is in line with the differences already observed in the literature between the mathematical-intensive fields of STEM and the other fields (Biology, psychology). However, this impact of gender at the individual level does not seem to create an aggregate impact: the share of women that enroll in the school is not sensible to the presence or the absence of options inside one track. Nevertheless, the distribution of women across majors inside the school seems to be related to the choice of options: the presence of options may therefore reinforce the role of gender norms that are already influencing the choice of majors for students. In addition, our survey distributed to candidates shows that candidates are partly unaware of the fact that options do not condition the choice of major once inside the school: imperfect information about exact exams conditions may then explain our results.

These results may have some implication for the organization of the ENS' entrance exams; as said before, it is unlikely that suppressing the options would increase significantly the proportion of women recruited, but it may help limit the imperfect information that students get on the exam organization. In fact, despite clear communication on the school website, a majority of the respondents think that their option choice will impact their choice of major. More generally, our results show that the settings of an exam are likely to become an artifact that impact outcomes they were not intended to impact. As a result, the settings of all exams should be scrutinized and debated beforehand, in order to keep only the settings that are necessary for adequate selection and evaluation and, therefore, limit the possible unintended consequences.

A certain amount of potential avenues are left open for future research. Firstly, once the data of this year exam will be available for research purpose, linking the survey answers to the data will offer a more detailed comprehension of the constitutive factors for the choices. In particular, it will allow a better understanding of whether the students who better understand the rules of the exam are more successful. Secondly, to pursue the study of the role of information and norms in academic success, one could also replicate this study with the sociological background information of the candidates. In particular, and since the question of equity in the access of the "Grandes Ecoles" is often debated, a similar methodology will help determine whether choices made by candidates are determined by their CPGE or by the wealth of their families. Finally, a study generalized on all the entrance tracks will allow the evaluation of the link between positive discrimination and gendered choices of option, which is suggested in our theoretical model but not empirically demonstrated, even though our results are compatible with the existence of such causality.

References

Akerlof, G. A., & Kranton, R. E. (2010, January 21). *Identity economics: How our identities shape our work, wages, and well-being* [Google-Books-ID: 1gKUpITuf00C]. Princeton University Press.

Anderson, L., Fryer, R., & Holt, C. (2005). Discrimination: Evidence from psychology and economics experiments. *Handbook on the Economics and Discrimination.*

Becker, G., Hubbard, W. H. J., & Murphy, K. (2010). Explaining the worldwide boom in higher education of women [Publisher: University of Chicago Press]. *Journal of Human Capital*, *4*(3), 203–241. Retrieved August 15, 2022, from https://econpapers.repec.org/article/ucpjhucap/doi_3a10.1086_2f657914.htm

Beede, D. N., Julian, T. A., Langdon, D., McKittrick, G., Khan, B., & Doms, M. E. (2011, August 1). Women in STEM: A gender gap to innovation. https://doi.org/10.2139/ssrn.1964782

Bloodhart, B., Balgopal, M. M., Casper, A. M. A., McMeeking, L. B. S., & Fischer, E. V. (2020). Outperforming yet undervalued: Undergraduate women in STEM [Publisher: Public Library of Science]. *PLOS ONE*, *15*(6), e0234685. https://doi.org/10.1371/journal.pone.0234685

Bonneau, C., Charousset, P., Grenet, J., & Thebeaut, G. (2021). Quelle démocratisation des grandes écoles depuis le milieu des années 2000? *IPP*, (30), 308.

Bouffard, T., Vezeau, C., & Simard, G. (2006). Motivations pour apprendre à l'école primaire : différences entre garçons et filles et selon les matières [Place: Paris cedex 14 Publisher: Presses Universitaires de France]. *Enfance*, *58*(4), 395–409. https://doi.org/10.3917/enf.584.0395

Breda, T., Jouini, E., Napp, C., & Thebault, G. (2020). Gender stereotypes can explain the gender-equality paradox [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *117*(49), 31063–31069. https://doi.org/10.1073/pnas.2008704117

Breda, T., & Ly, S. T. (2015). Professors in core science fields are not always biased against women:evidence from france. *American Economic Journal: Applied Economics*, *7*(4), 53–75. https://doi.org/http://dx.doi.org/10.1257/app.20140022

Breda, T., & Napp, C. (2019). Girls' comparative advantage in reading can largely explain the gender gap in math-related fields [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *116*(31), 15435–15440. https://doi.org/10.1073/pnas.1905779116

Brenøe, A. A., & Zölitz, U. (2020). Exposure to more female peers widens the gender gap in STEM participation [Publisher: The University of Chicago Press]. *Journal of Labor Economics*, *38*(4), 1009–1054. https://doi.org/10.1086/706646

Calvez, C., & Viry, S. (2018). *Rapport d'information déposé par la délégation de l'Assemblée nationale aux droits des femmes et à l'égalité des chances entre les hommes et les femmes sur les femmes et les sciences* (Rapport d'information No. 1016). Assemblée Nationale. Retrieved August 28, 2022, from https://www.assemblee-nationale.fr/dyn/15/rapports/ega/l15b1016_rapport-information

Card, D., & Payne, A. A. (2017, September). High school choices and the gender gap in STEM. https://doi.org/10.3386/w23769

Carlana, M. (2019). Implicit stereotypes: Evidence from teachers' gender bias*. *The Quarterly Journal of Economics*, *134*(3), 1163–1224. https://doi.org/10.1093/qje/qjz008

Carrell, S. E., Page, M. E., & West, J. E. (2010). Sex and science: How professor gender perpetuates the gender gap*. *The Quarterly Journal of Economics*, *125*(3), 1101–1144. https://doi.org/10.1162/qjec.2010.125.3.1101

Ceci, S. J., Ginther, D. K., Kahn, S., & Williams, W. M. (2014). Women in academic science: A changing landscape [Publisher: SAGE Publications Inc]. *Psychological Science in the Public Interest*, *15*(3), 75–141. https://doi.org/10.1177/1529100614541236

Ceci, S. J., & Williams, W. M. (2009, September 2). *The mathematics of sex: How biology and society conspire to limit talented women and girls* [Google-Books-ID: FiwGfZ6yUQIC]. Oxford University Press.

Ceci, S. J., & Williams, W. M. (2010). Sex differences in math-intensive fields [Publisher: SAGE Publications Inc]. *Current Directions in Psychological Science*, *19*(5), 275–279. https://doi.org/10.1177/0963721410383241

Chabanon, L., & Jouvenceau, M. (2022). Femmes et hommes, l'égalité en question. *INSEE Références*. Retrieved August 14, 2022, from https://www.insee.fr/fr/statistiques/6047791?sommaire=6047805

Chabanon, L., & Steinmetz, C. (2018). ÉCARTS DE PERFORMANCES DES ÉLÈVES SELON LE SEXE. *ÉDUCATION & FORMATIONS*, (96), 20.

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin*, *143*(1), 1–35. https://doi.org/10.1037/bul0000052

Delaney, J., & Devereux, P. (2021). High school rank in math and english and the gender gap in STEM [Publisher: Elsevier]. *Labour Economics*, *69*. Retrieved August 16, 2022, from https://econpapers.repec.org/article/eeelabeco/v_3a69_3ay_3a2021_3ai_3ac_3as092753712100004x.htm

Delaney, J., & Devereux, P. J. (2021). Gender and educational achievement: Stylized facts and causal evidence. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3775979

Delaney, J. M., & Devereux, P. J. (2019). Understanding gender differences in STEM: Evidence from college applications. *Economics of Education Review*, *72*, 219–238. https://doi.org/10.1016/j.econedurev.2019.06.002

Diprete, T. A., & Buchmann, C. (2006). Gender-specific trends in the value of education and the emerging gender gap in college completion. *Demography*, *43*(1), 1–24. https://doi.org/10.1353/dem.2006.0003

Fryer, R. J. (2007). Belief flipping in a dynamic model of statistical discrimination [Publisher: Elsevier]. *Journal of Public Economics*, *91*(5), 1151–1166. Retrieved August 16, 2022, from https://ideas.repec.org/a/eee/pubeco/v91y2007i5-6p1151-1166.html

Ginther, D. K., & Kahn, S. (2006, November). Does science promote women? evidence from academia 1973-2001. https://doi.org/10.3386/w12691

Grenet, J., Bechichi, N., & Thebault, G. (2021). D'admission post-bac à parcoursup : Quels effets sur la répartition des néo-bacheliers dans les formations d'enseignement supérieur ? *France, Portrait Social*. https://www.insee.fr/fr/statistiques/5432519?sommaire=5435421

Hlavac, M. (n.d.). Stargazer: Well-formatted regression and summary statistics tables. r package version 5.2.3. https://doi.org/https://CRAN.R-project.org/package=stargazer

Kahn, S., & Ginther, D. K. (2017, June 1). Women and stem. Retrieved August 14, 2022, from https://papers.ssrn.com/abstract=2988746

Kiefer, A. K., & Sekaquaptewa, D. (2007). Implicit stereotypes and women's math performance: How implicit gender-math stereotypes influence women's susceptibility to stereotype threat. *Journal of Experimental Social Psychology*, *43*(5), 825–832. https://doi.org/10.1016/j.jesp.2006.08.004

Marks, G. N. (2008). Accounting for the gender gaps in student performance in reading and mathematics: Evidence from 31 countries. *Oxford Review of Education*, *34*(1), 89–109. https://doi.org/10.1080/03054980701565279

O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance [Publisher: SAGE Publications Inc]. *Personality and Social Psychology Bulletin*, *29*(6), 782–789. https://doi.org/10.1177/0146167203029006010

OECD. (2021). *Education at a Glance 2021: OECD Indicators*. Organisation for Economic Co-operation; Development. Retrieved August 14, 2022, from https://www.oecd-ilibrary.org/fr/education/education-at-a-glance-2021_b35a14e5-en

Perez-Felkner, L., Nix, S., & Thomas, K. (2017). Gendered pathways: How mathematics ability beliefs shape secondary and postsecondary course and degree field choices. *Frontiers in Psychology*, *8*, 386. https://doi.org/10.3389/fpsyg.2017.00386

Shapiro, C. A., & Sax, L. J. (2011). Major selection and persistence for women in STEM [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ir.404]. *New Directions for Institutional Research*, *2011*(152), 5–18. https://doi.org/10.1002/ir.404

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance [Place: Netherlands Publisher: Elsevier Science]. *Journal of Experimental Social Psychology*, *35*(1), 4–28. https://doi.org/10.1006/jesp.1998.1373

Stoet, G., & Geary, D. C. (2012). Can stereotype threat explain the gender gap in mathematics performance and achievement? [Publisher: SAGE Publications Inc]. *Review of General Psychology*, *16*(1), 93–102. https://doi.org/10.1037/a0026617

Stoet, G., & Geary, D. C. (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological Science*, *29*(4), 581–593. https://doi.org/10.1177/0956797617741719

Trapani, J., & Hale, K. (2022, February 22). *Higher education in science and engineering*. National Science Foundation. Retrieved August 15, 2022, from https://ncses.nsf.gov/pubs/nsb20223/demographic-attributes-of-s-e-degree-recipients

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance [Place: US Publisher: American Psychological Association]. *Journal of Educational Psychology*, *101*(4), 817–835. https://doi.org/10.1037/a0016127

Weinberger, C. J. (1999). Mathematical college majors and the gender gap in wages. *Industrial Relations: A Journal of Economy and Society*, *38*(3), 407–413. https://doi.org/10.1111/0019-8676.00134

Xie, Y., & Shauman, K. A. (2003). *Women in science*. Harvard university press, retrieved August 15, 2022, from https://elibrary.ru/item.asp?id=19978280

A Multiple-choice survey

A.1 Flyer distributed to candidates

Questionnaire pour les admissibles au concours - SCIENCE

Afin de réaliser une enquête (dans le cadre d'un mémoire de master) sur les candidats et candidates au concours de l'ENS, nous vous mettons à disposition un questionnaire. Répondre à ce questionnaire ne prend que cinq minutes, et nous aiderait beaucoup ! Vos réponses seront utilisées exclusivement à des fins de recherche. Elles seront traitées de façon anonyme dans le respect des règles de confidentialité en vigueur. En particulier, aucun personnel administratif de l'école ni membre du jury n'aura connaissance de vos réponses.

Vous pouvez nous laisser vos coordonnées afin de recevoir le questionnaire par email (voir feuille prévue à cet effet) ou accéder au questionnaire par le lien suivant : https://fr.surveymonkey.com/r/V9VZST7 ou via le QR code ci-dessous :

Merci d'avance !

Figure 17: Flyer distributed to incite eligible candidates to fulfill a survey

A.2 Content of the survey

Concours et choix d'option - séries scientifiques

Informations préalables

Ces données sont collectées uniquement dans une finalité de recherche et seront anonymisées avant tout traitement statistique. Uniquement si vous l'acceptez à la fin de ce questionnaire, elles seront également appariées à vos résultats au concours sur la base de vos nom et prénom, puis à nouveau anonymisées avant traitement. Elles seront dans tous les cas conservées au plus dix ans. Vous gardez par ailleurs à tout moment un droit d'accès, de modification et de suppression à vos données personnelles ; pour exercer ce droit, contactez clemence.gardette@ens.fr.

L'Ecole Normale Supérieure de Paris ne saurait être tenue responsable de ce questionnaire, ou de l'usage qui en est fait ; de façon similaire, aucune réponse de ce questionnaire ne sera connue par l'école, les jurys ou des professeurs de l'ENS.

Concours et choix d'option - séries scientifiques

Informations biographiques

* 1. À quel genre vous identifiez-vous ?

- ◯ Femme
- ◯ Homme
- ◯ Non-binaire
- ◯ Préfère ne pas répondre

* 2. Etes-vous étudiant.e boursier ?

- ◯ Oui, boursier CROUS
- ◯ Oui, mais non boursier CROUS
- ◯ Non
- ◯ Préfère ne pas répondre

3. Combien de fois avez-vous passé les concours de l'ENS ? (en incluant 2022)

- ◯ Une fois
- ◯ Deux fois
- ◯ Trois fois ou plus

4. Où votre classe préparatoire est-elle située ?

- ◯ Paris
- ◯ Banlieue parisienne
- ◯ Province
- ◯ Aucun des éléments ci-dessus

Si vous le souhaitez, vous pouvez mettre le nom de votre prépa ici

[ ]

* 5. A quel concours êtes vous inscrits ?

- ◯ MPI
- ◯ Info
- ◯ PC
- ◯ BCPST
- ◯ PSI
- ◯ Autre

Concours et choix d'option - séries scientifiques

Choix d'option - MPI

* 6. Quelle option avez-vous choisie lors de votre passage du concours en 2022 ?

- ◯ Maths-physique
- ◯ Maths-physique-info

7. Quelle est la date (jour et heure) de votre convocation à l'oral correspondant à votre option (physique ou informatique) ?

Date/Heure

Date	Heure		AM/PM
JJ/MM/AAAA	hh	mm	-

8. Quelle est la date (jour et heure) de votre convocation à l'oral de mathématiques I ?

Date/Heure

Date	Heure		AM/PM
JJ/MM/AAAA	hh	mm	-

Concours et choix d'option - séries scientifiques

Choix d'option - PC

* 9. Quelle option avez-vous choisie lors votre passage du concours en 2022 ?

- Physique
- Chimie

* 10. Quelle est la date (jour et heure) de votre convocation à l'oral de la matière que vous avez choisi en option ?

Date/Heure

Date JJ/MM/AAAA | Heure hh mm | AM/PM -

* 11. Quelle est la date (jour et heure) de votre convocation à l'oral de mathématiques ?

Date/Heure

Date JJ/MM/AAAA | Heure hh mm | AM/PM -

Concours et choix d'option - séries scientifiques

Choix d'option - BCPST

* 12. Quelle option avez-vous choisie lors de votre passage du concours en 2022 ?

- Biologie
- Sciences de la terre

13. Quelle est la date (jour et heure) de votre convocation à l'oral de la matière que vous avez choisi en option ?

Date/Heure

Date JJ/MM/AAAA | Heure hh mm | AM/PM -

14. Quelle est la date (jour et heure) de votre convocation à l'oral de physique ?

Date/Heure

Date JJ/MM/AAAA | Heure hh mm | AM/PM -

Concours et choix d'option - séries scientifiques

Choix d'option - suite

15. Comment avez-vous choisi votre option ? (plusieurs réponses possibles)

- [ ] C'est la discipline qui m'intéressait le plus
- [ ] C'est la discipline dans laquelle je voulais obtenir une équivalence
- [ ] C'est la discipline dans laquelle je considérais être le/la meilleure en CPGE
- [ ] Autre (veuillez préciser)
- [ ] C'est la discipline dans laquelle je souhaite poursuivre mes études
- [ ] Je n'ai pas accordé d'importance à l'impact du choix de ma spécialité sur les concours de l'ENS

16. Plus spécifiquement, diriez-vous que vous avez choisi votre spécialité pour maximiser vos chances d'intégrer l'ENS ?

- ( ) Oui
- ( ) Non

17. De façon purement subjective, à combien évaluriez-vous vos chances d'être admis.e à l'ENS en 2022 ?

☆ ☆ ☆ ☆ ☆

18. Qui vous a accompagné pour faire votre choix de spécialité en prépa ? (plusieurs réponses possibles)

- [ ] Mes pairs de classe préparatoire
- [ ] Mes enseignants de classe préparatoire
- [ ] Ma famille
- [ ] Mes amis (hors classe préparatoire)
- [ ] Personne, j'ai fait mon choix seul sans accompagnement
- [ ] Autre (veuillez préciser)

19. Si vous pouviez changer de spécialité, le feriez-vous ?

- ( ) Oui
- ( ) Peut-être
- ( ) Je ne sais pas
- ( ) Probablement pas
- ( ) Non

Si vous avez répondu "oui" ou "peut-être", quelle spécialité auriez-vous préféré prendre ?

20. Même s'il est très difficile de savoir, quelle note pensez-vous obtenir à votre épreuve de spécialité ?

0 10 20

21. Votre professeur d'option en 2021-2022 était...

- Un homme
- Une femme
- Autre/préfère ne pas répondre

22. D'après vous, la discipline que vous avez choisie comme spécialité est une discipline stéréotypée...

- Très masculine
- Plutôt masculine
- Non associée à un genre particulier
- Plutôt féminine
- Très féminine
- Autre (veuillez préciser)

Concours et choix d'option - séries scientifiques

Scolarité envisagée à l'ENS

23. Selon vous, est-ce que votre choix de spécialité durant le concours conditionnera, si vous êtes admis.e, votre scolarité à l'ENS ?

- Oui, totalement
- Oui, en partie
- Je ne sais pas
- Probablement pas
- Non, pas du tout

* 24. Quel(s) département(s) seriez-vous susceptible de rejoindre à Ulm en cas d'admission ? (plusieurs réponses possibles)

- Etudes cognitives
- Informatique
- Biologie
- Chimie
- Géosciences
- Mathématiques et applications
- Physique
- Environnement et société
- Je ne sais pas
- Autre (département littéraire)

Concours et choix d'option - séries scientifiques

Concours et conditions matérielles

25. Dans quelles conditions matérielles avez-vous passé vos oraux ?

★ ★ ★ ★ ★

26. Si vous avez eu des frais dus au passage des concours, à combien les évaluriez-vous ?

- ○ Moins de 100 euros
- ○ Entre 100 et 200 euros
- ○ Entre 200 et 300 euros
- ○ Plus de 300 euros
- ○ Je ne sais pas

27. Quels coûts supplémentaires ont été engendré par le passage des oraux ?

- ☐ Frais d'hébergement
- ☐ Frais de transports
- ☐ Frais divers
- ☐ Aucun des éléments ci-dessus

Si vous avez répondu frais divers, vous pouvez précisez ici.

28. Saviez-vous que vous pouviez être hébergé par l'ENS durant vos oraux ?

- ○ Oui
- ○ Non

Concours et choix d'option - séries scientifiques

Logement - non informé

29. Si vous aviez su que l'ENS proposait des logements durant le passage des oraux, auriez-vous souhaité en bénéficier ?

- ○ Oui
- ○ Non

Concours et choix d'option - séries scientifiques

Logement - informé

30. Avez-vous été logé à l'ENS durant le passage de vos oraux ?

- ◯ Oui
- ◯ Non

Concours et choix d'option - séries scientifiques

Fin du sondage

31. Votre nom :

[ ]

32. Votre prénom :

[ ]

* 33. J'accepte que mes réponses soient appariées à mes résultats au concours avant d'être anonymisées.

- ☐ Oui
- ☐ Non

A.3 Display of all the answers to the survey

Questionnaire admissibles 2022

Clémence Gardette

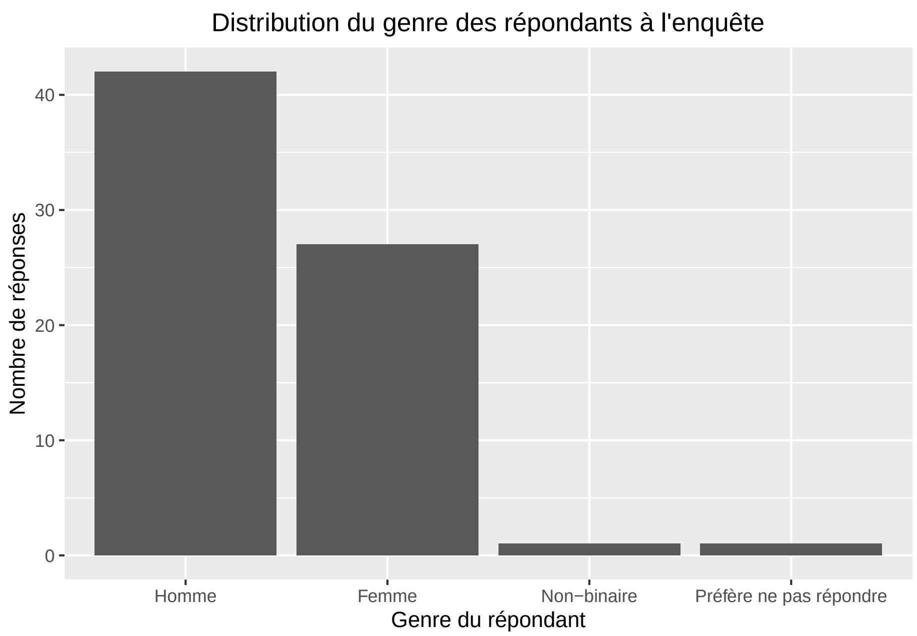
13 juillet 2022

Compte-rendu du questionnaire distribué aux admissibles sciences

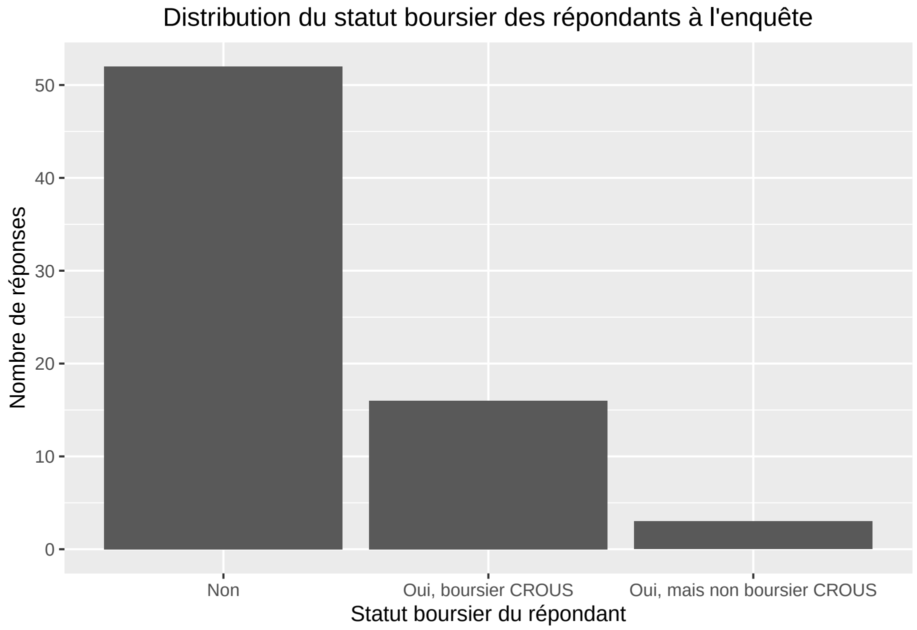
Ce document a pour objectif de proposer un premier compte-rendu des résultats obtenus en réponse au questionnaire distribué à une partie des admissibles aux concours science 2022. Ce questionnaire a été distribué par le biais d'un flyer, qui contenait un QR code renvoyant audit questionnaire administré en ligne. Ce flyer était disponible en salle Beckett (salle des admisseurs), du 28 juin au 8 juillet 2022. Environ 150 flyers ont été imprimés, et 71 réponses ont été recueillies. Ce compte-rendu est divisé en sept parties : la première partie contient les statistiques descriptives caractérisant les répondants (genre, origine géographique, filière. . .). Les trois parties suivantes se focalisent chacune sur une filière : MPI, PC et BCPST. Les deux parties suivantes cherchent à étudier les modalités du choix d'option parmi les candidats, ainsi que la scolarité qu'ils envisagent de suivre une fois à l'ENS. Enfin, la dernière partie présente les conditions matérielles dans lesquels les candidats rapportent avoir passés leurs concours.

Questions générales

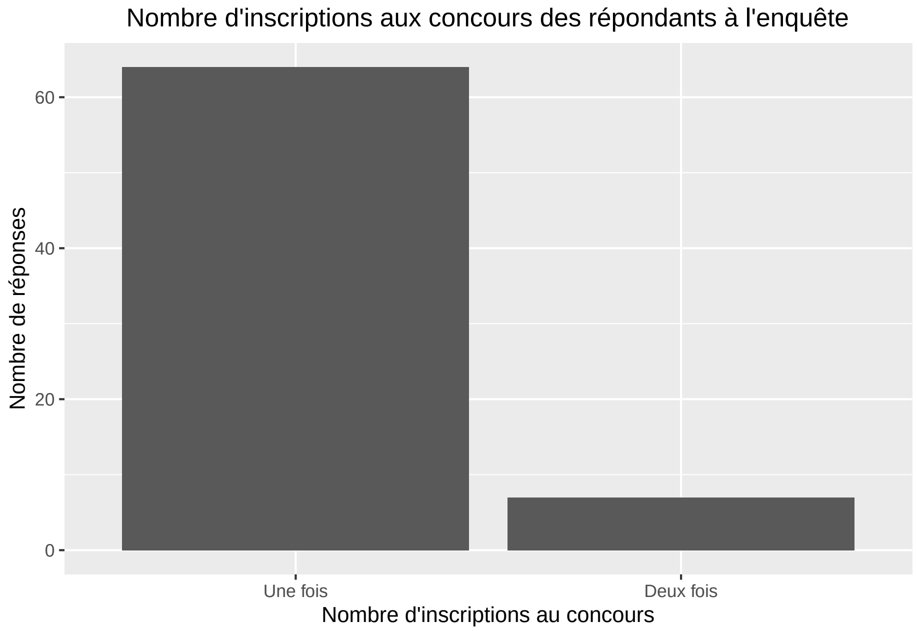
Q1 : A quel genre vous identifiez-vous ?



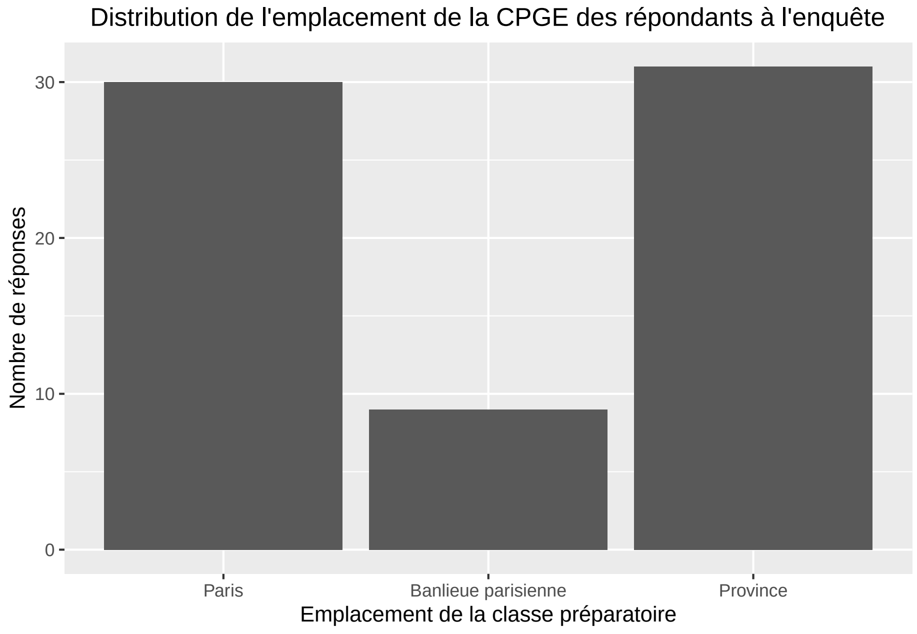
Q2 : Êtes-vous étudiant.e boursier.e ?



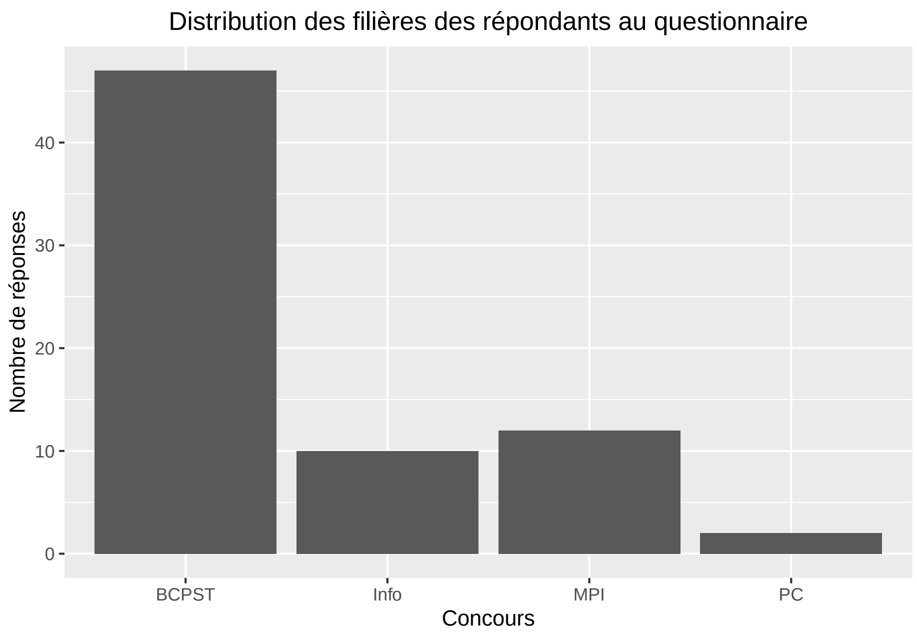
Q3 : Combien de fois avez-vous passé les concours de l'ENS ? (en incluant 2022)



Q4 : Où votre classe préparatoire est-elle située ?

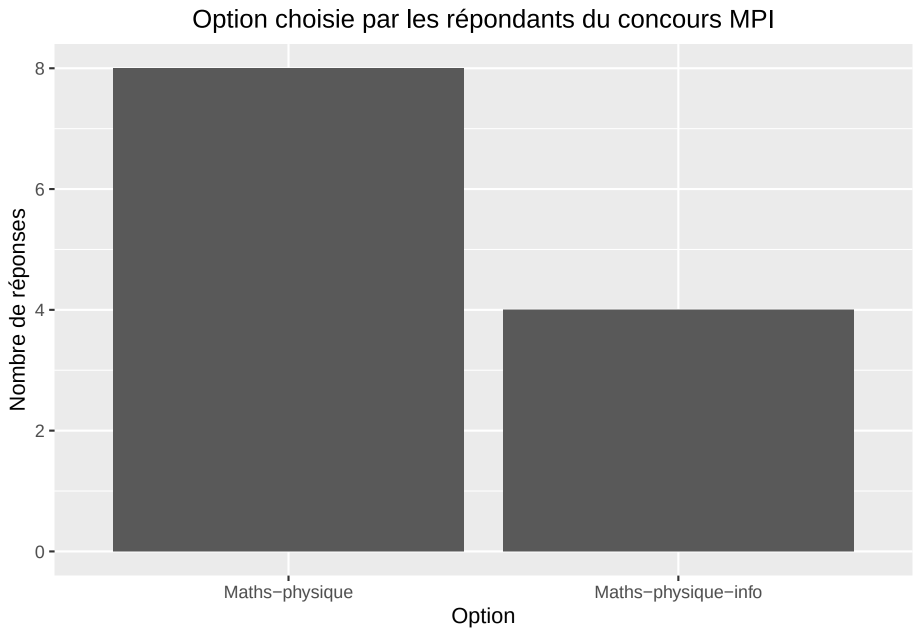


Q5 : A quel concours êtes vous inscrits ?

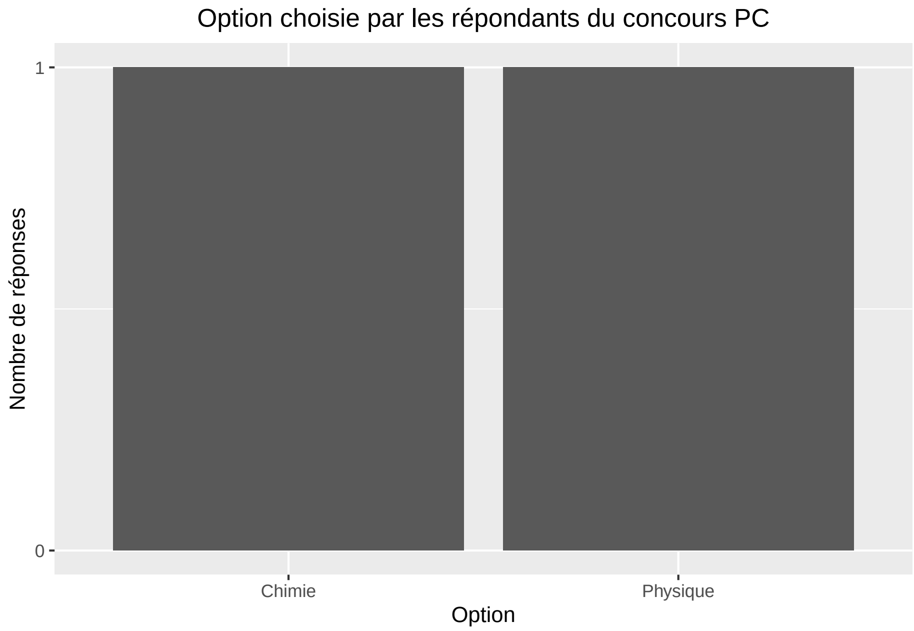


Sur ce graphe, on observe une très forte surreprésentation des répondants en BCPST par rapport aux autres filières, sans mesure avec le nombre d'inscrits dans chaque filière. Cette surreprésentation est causée par le mode d'administration du questionnaire : en effet, ce dernier est distribué en salle des admisseurs. Or, l'organisation des concours oblige les BCPST a passer systématiquement en salle des admisseurs avant leurs oraux, alors que ce n'est pas le cas pour les autres filières scientifiques : les candidats des filières MPI, Info et PC sont donc minoritaires à recevoir le questionnaire, ce qui explique la différence de réponses. Les analyses menées ci-dessous sont donc particulièrement pertinentes pour les BCPST, mais sont à prendre avec précaution pour les autres filières, dans lesquelles le nombre de répondants est beaucoup plus faible.

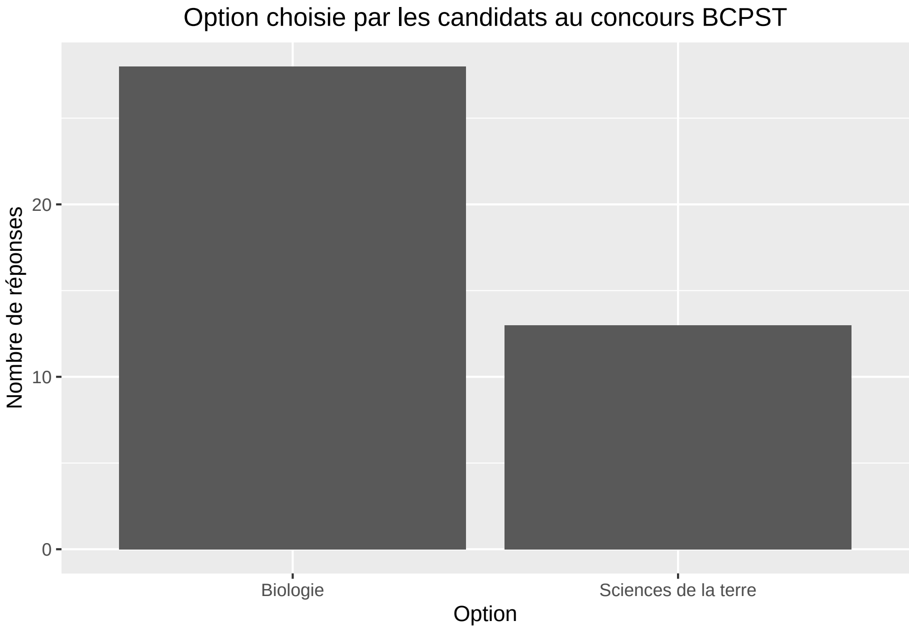
Choix d'option - MPI



Choix d'option - PC



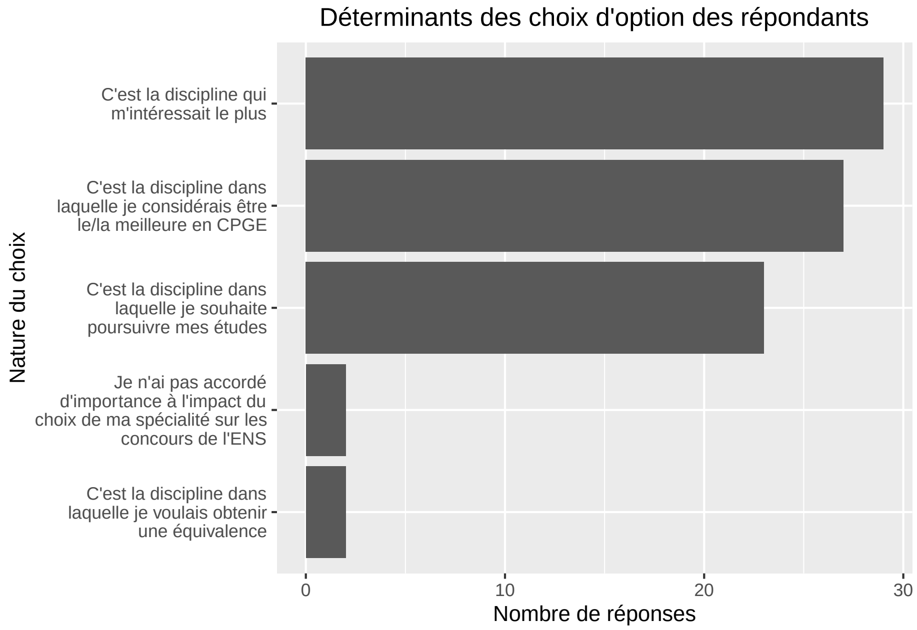
Choix d'option - BCPST



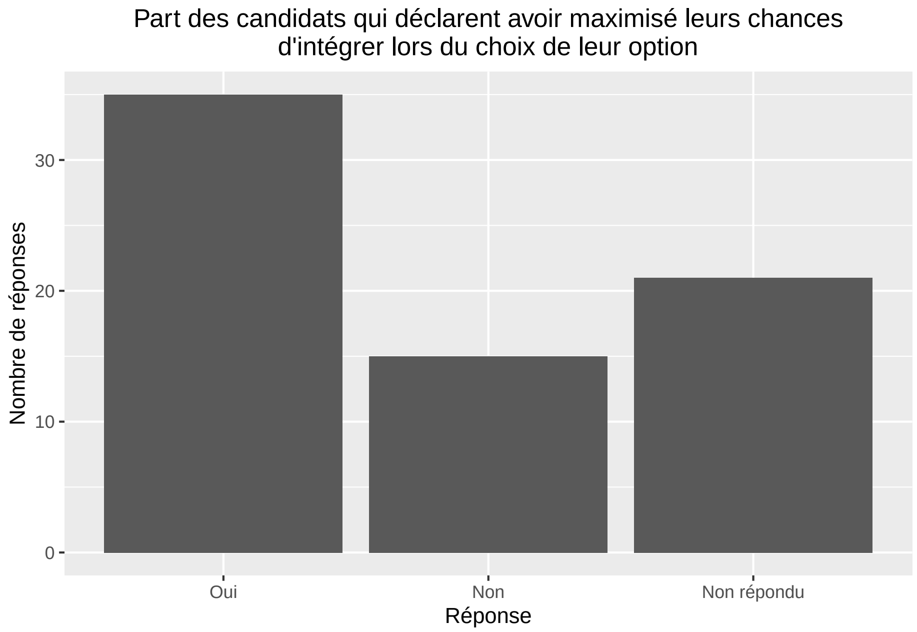
Etude du choix de leur option par les candidats

Cette sous-partie cherche à éclairer la manière dont les candidats choisissent leur option au concours de l'ENS : ont-ils compris les implications que ce choix aura dans leur scolarité ? Font-ils ce choix seuls, ou sont-ils aidés par leurs professeurs ? Ces questions cherchent, dans le cadre particulier d'un mémoire d'économie, à comprendre si les candidats se comportent comme des “agents rationnels”, qui cherchent à optimiser leurs chances d'intégrer, ou si d'autres facteurs rentrent en jeu. Ces facteurs peuvent être conscients (par exemple, les candidats peuvent choisir volontairement la matière qu'ils préfèrent, même si ce n'est pas celle dans laquelle ils sont les meilleurs) ou inconscients (les candidats peuvent être influencés par leur milieu social, leur professeur, etc.)

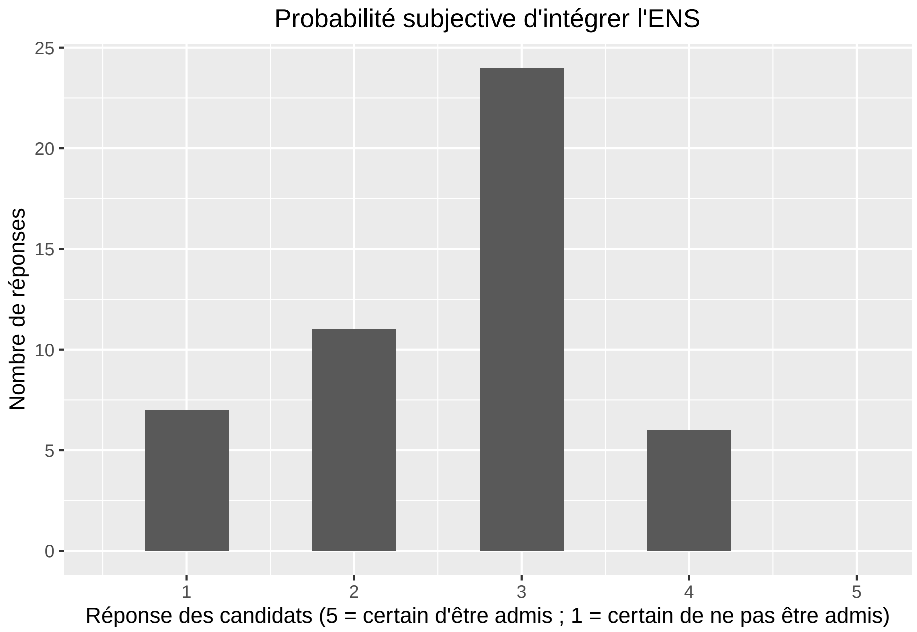
Q15 : Comment avez-vous choisi votre option ?



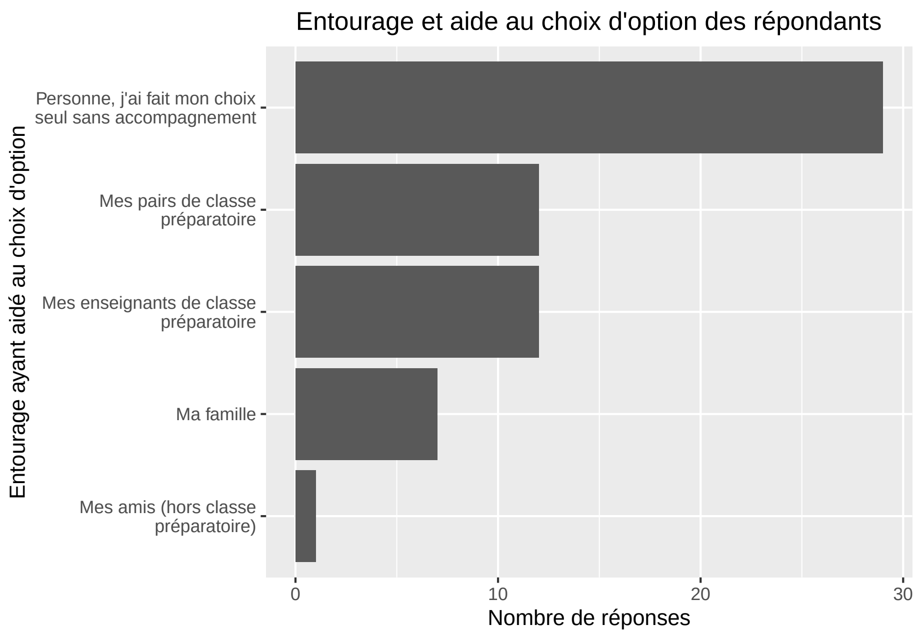
Q16 : Plus spécifiquement, diriez-vous que vous avez choisi votre spécialité pour maximiser vos chances d'intégrer l'ENS ?



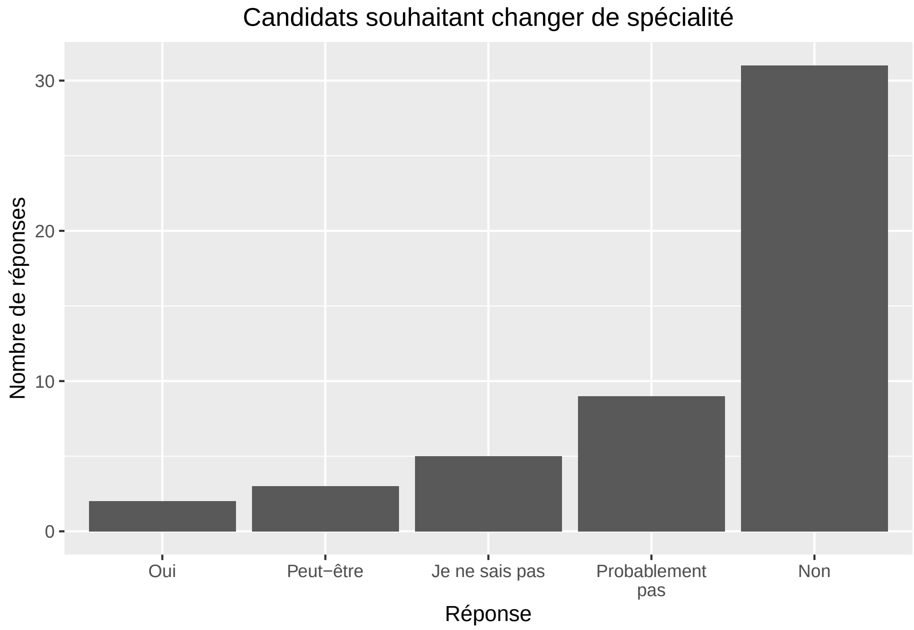
Q17 : De façon purement subjective, à combien évaluriez-vous vos chances d'être admis.e à l'ENS en 2022 ?



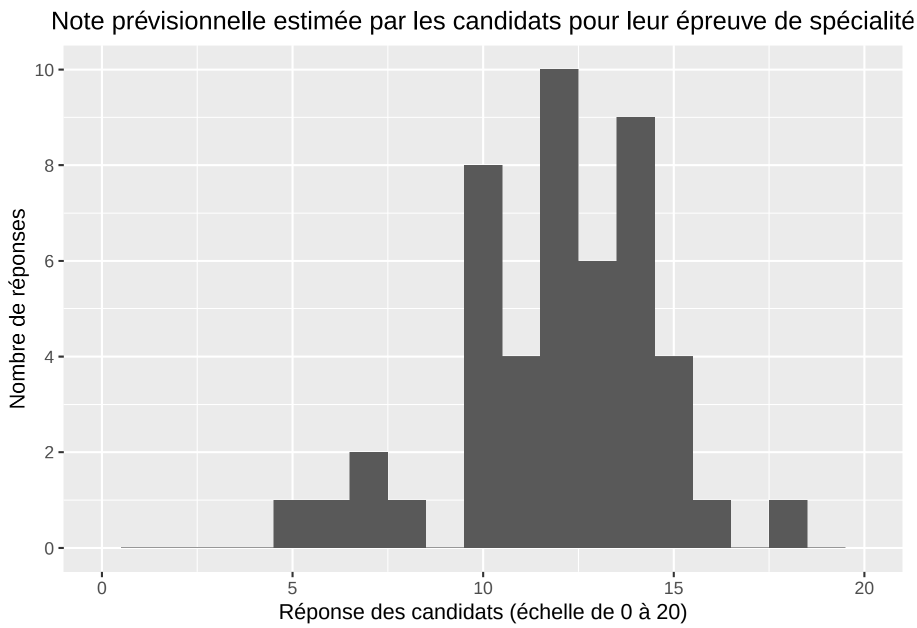
Q18 : Qui vous a accompagné pour faire votre choix de spécialité en prépa ? (plusieurs réponses possibles)



Q19 : Si vous pouviez changer de spécialité, le feriez-vous ?

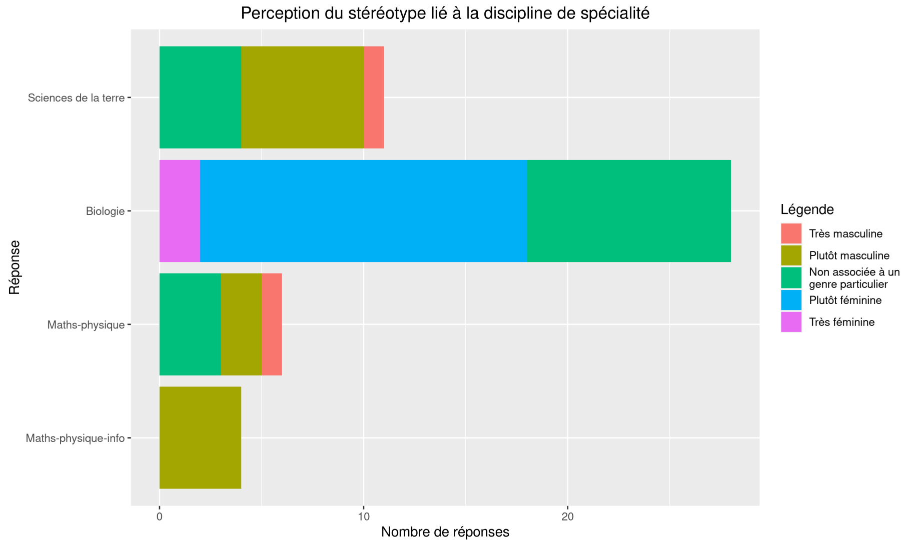


Q20 : Même s'il est très difficile de savoir, quelle note pensez-vous obtenir à votre épreuve de spécialité ?



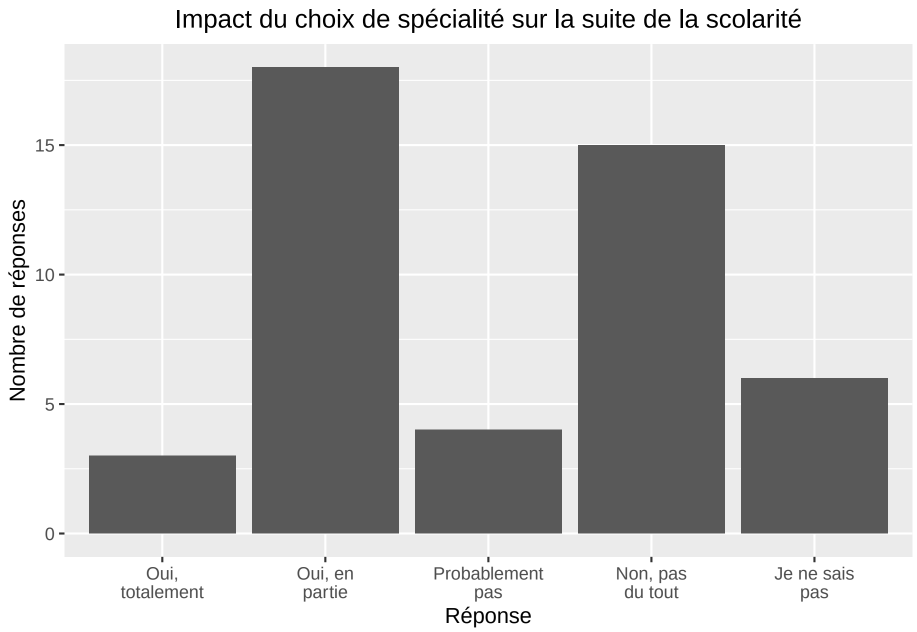
Q21 : Votre professeur d'option en 2021-2022 était. . .

Q22 : D'après vous, la discipline que vous avez choisie comme spécialité est une discipline stéréotypée. . .

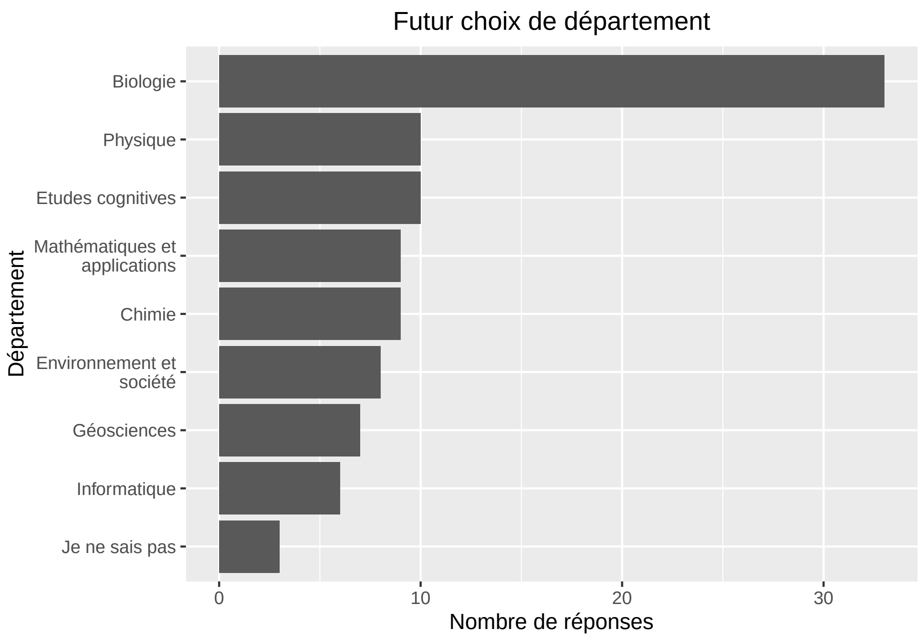


Scolarité envisagée à l'ENS

Q23 : Selon vous, est-ce que votre choix de spécialité durant le concours conditionnera, si vous êtes admis.e, votre scolarité à l'ENS ?

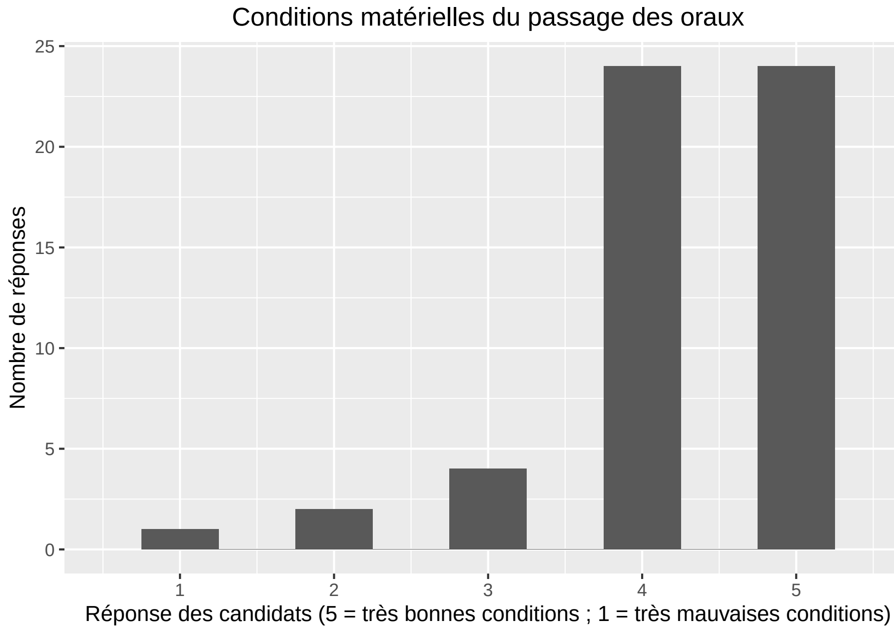


Q24 : Quel(s) département(s) seriez-vous susceptible de rejoindre à Ulm en cas d'admission ? (plusieurs réponses possibles)

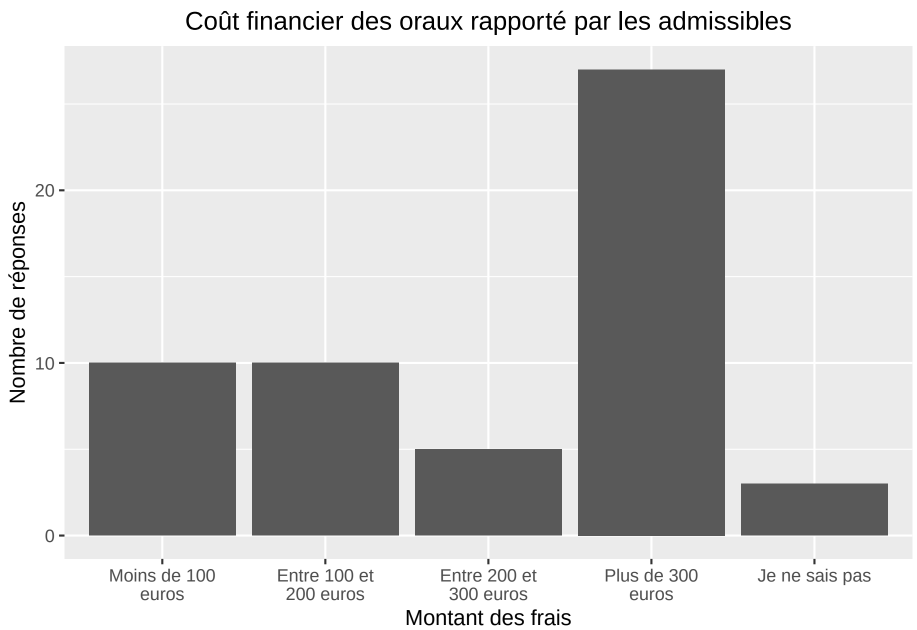


Concours et conditions matérielles

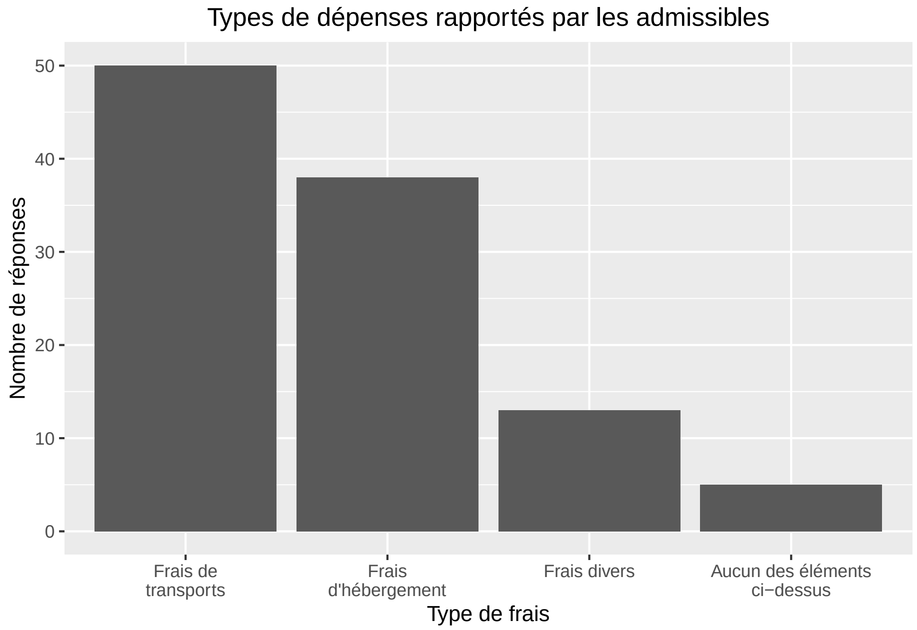
Q25 : Dans quelles conditions matérielles avez-vous passé vos oraux ?



Q26 : Si vous avez eu des frais dus au passage des concours, à combien les évaluriez-vous ?

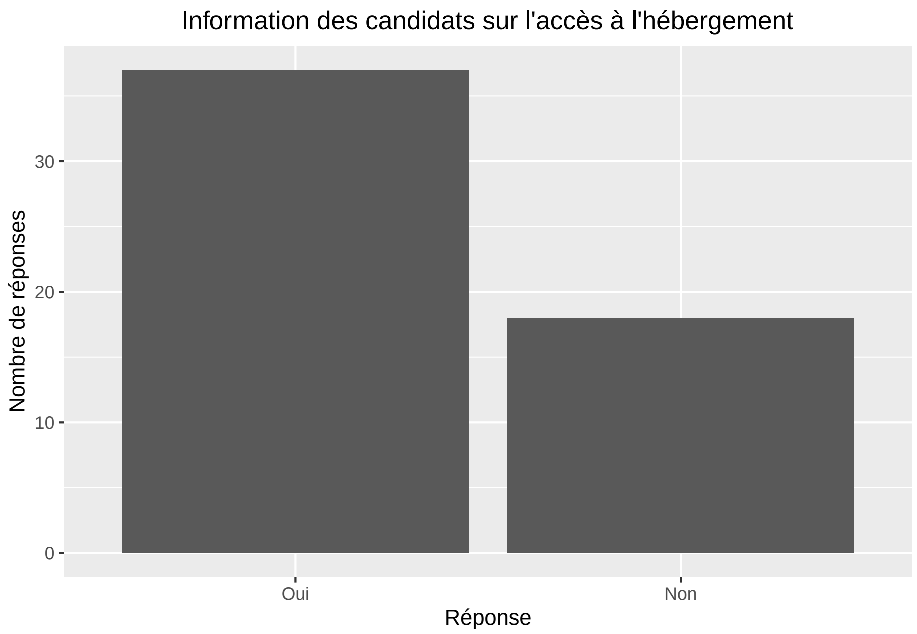


Q27 : Quels coûts supplémentaires ont été engendré par le passage des oraux ?

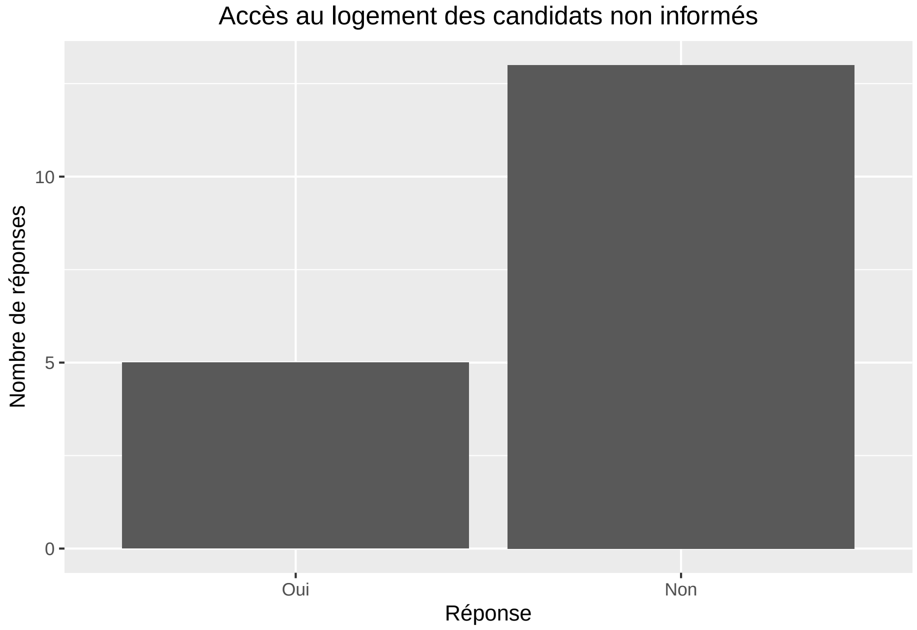


Les répondants pouvaient également ce qu'ils entendaient par frais divers ; la majorité d'entre eux ayant choisi de préciser indiquent qu'ils ont du payer des sommes non négligeables en frais d'alimentation. En effet, les candidats indiquent devoir se nourrir au restaurant ou de nourriture à emporter, ce qui reste sensiblement plus cher que le coût du self durant l'année de préparation.

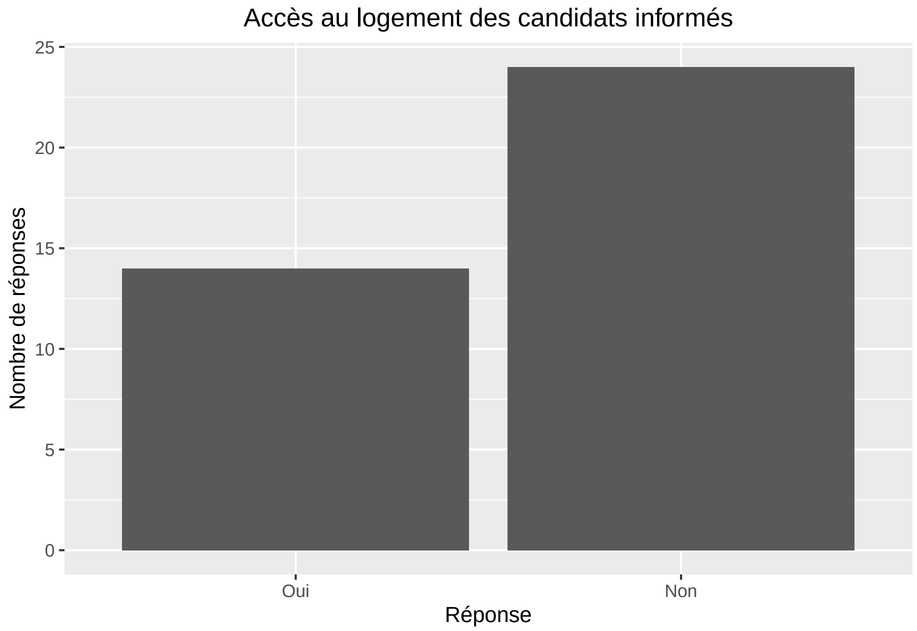
Q28 : Saviez-vous que vous pouviez être hébergé par l'ENS durant vos oraux ?



Q29 : Si vous aviez su que l'ENS proposait des logements durant le passage des oraux, auriez-vous souhaité en bénéficier ? (candidats ayant répondu "Non" à la question 28)



Q30 : Avez-vous été logé à l'ENS durant le passage de vos oraux ? (candidats ayant répondu "Oui" à la question 28)



B Computations of the theoretical model

B.1 Analytical solution of the offer side

To solve this model analytically, we need only specify a law for the n_m levels, which will create a law for the difference $\Delta = n_{\mathrm{B}} - n_{\mathrm{A}}$. Assuming that the levels are represented in points from 0 to 20, a natural specification would be a normal distribution, for example centered in 10. However, this specification would not allow computing analytically a solution, since one cannot express the density of the normal distribution. Therefore, an alternative is the use of a uniform distribution $U[0, 20]$ for the levels, which are thus fairly distributed in the population between 0 and 20. The sum of two uniform distributions being a triangular distribution, we have $\Delta = n_text B - n_{\mathrm{A}} \sim Triang(-20, 20, 0)$. Finally, although the distribution function of the triangular distribution is expressed on the left and on the right of 0, it is not necessary here to consider the right (positive) part, since we will consider that the difference $d_{\mathrm{B,A}} - \beta(r_{\mathrm{B}} - 0.5)$ is negative. This assumption amounts to saying that positive discrimination does not reverse the social sanction, i.e. that schools try to compensate for the distortion without completely reversing the norm. Thus, since the distribution function of a $Triang(a, b, c)$ law is

$$F(x) = \frac{(x-a)^2}{(b-a)(c-a)}$$

for $a < x < c$, we can compute the equilibrium ratio, which is the root between 0 and 1 implied by the distribution function. Analytically, this solution is equal to

$$r_{\mathrm{B}} = \frac{(800 + 2\beta X) - \sqrt{D}}{2\beta^2}$$

with the notations

$$\begin{cases} X = (0.5\beta - d_{\mathrm{B,A}} + 20) \\ D = (800 + 2X\beta)^2 - 4\beta^2 X^2 \end{cases}$$

C Additional regression tables

C.1 Choice of option with regards with gender, by track, with ranking instead of grades

Table 10: Prediction of option choice according to the gender and the grades

	Dependent variable:			
	Likelihood of choosing physics		Likelihood of choosing geology	
	(1)	(2)	(3)	(4)
Woman	-0.571^{***}	-0.224^{***}	-0.229^{**}	-0.017^{***}
	(0.083)	(0.032)	(0.092)	(0.006)
Rank in Chemistry	-1.530^{***}	-0.592^{***}		
	(0.129)	(0.050)		
Rank in Physics	1.369^{***}	0.530^{***}		
	(0.130)	(0.050)		
Rank in biology			0.399^{*}	0.030^{*}
			(0.205)	(0.015)
Rank in geology			0.134	0.010
			(0.201)	(0.015)
Constant	0.415^{***}		1.702^{***}	
	(0.089)		(0.099)	
Observations	1,470	1,470	3,037	3,037
Log Likelihood	-857.304	-857.304	-463.313	-463.313
Akaike Inf. Crit.	1,722.609	1,722.609	934.627	934.627
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 11: Prediction of option choice according to the gender and the grades - oral grades

	Dependent variable:			
	Likelihood of choosing physics		Likelihood of choosing geology	
	(1)	(2)	(3)	(4)
Woman	-0.398^{**}	-0.155^{**}	-0.087	-0.020
	(0.172)	(0.068)	(0.191)	(0.043)
Rank in Chemistry	-0.649^{***}	-0.245^{***}		
	(0.235)	(0.089)		
Rank in Physics	0.685^{***}	0.259^{***}		
	(0.230)	(0.087)		
Rank in biology			-0.332	-0.075
			(0.339)	(0.077)
Rank in geology			1.193^{***}	0.271^{***}
			(0.353)	(0.078)
Constant	0.385^{**}		-1.417^{***}	
	(0.159)		(0.258)	
Observations	402	402	268	268
Log Likelihood	-252.524	-252.524	-110.146	-110.146
Akaike Inf. Crit.	513.048	513.048	228.293	228.293

Note: *p<0.1; **p<0.05; ***p<0.01

C.2 Choice of option with regards with gender, written grades, for eligible candidates only

Table 12: Prediction of option choice according to the gender and the grades - eligible candidates only

	Dependent variable:			
	Likelihood of choosing physics		Likelihood of choosing geology	
	(1)	(2)	(3)	(4)
Woman	-0.577^{***}	-0.224^{***}	0.016	0.003
	(0.170)	(0.066)	(0.189)	(0.040)
Grade in Chemistry	-0.149^{***}	-0.056^{***}		
	(0.028)	(0.010)		
Grade in Physics	0.105^{***}	0.039^{***}		
	(0.024)	(0.009)		
Grade in biology			-0.109^{***}	-0.023^{***}
			(0.037)	(0.008)
Grade in geology			0.115^{***}	0.024^{***}
			(0.031)	(0.006)
Constant	1.027^{**}		-1.149^{*}	
	(0.455)		(0.649)	
Observations	412	412	291	291
Log Likelihood	-245.030	-245.030	-112.752	-112.752
Akaike Inf. Crit.	498.061	498.061	233.503	233.503
Note:			*p<0.1; **p<0.05; ***p<0.01	

C.3 Impact of option choice on the exit department

Table 13: Prediction of department choice

	Dependent variable: exit department							
	Likelihood of choosing physics as a major				Likelihood of choosing geology as a major			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Woman	0.083	0.0002	0.216	0.059	−0.470	−0.062	−0.102	−0.019
	(0.692)	(0.122)	(0.529)	(0.135)	(0.506)	(0.069)	(0.396)	(0.075)
Grade in Chemistry	−0.062	−0.0001	−0.130*	−0.038*				
	(0.085)	(0.100)	(0.070)	(0.020)				
Grade in Physics	0.047	0.0001	0.048	0.014				
	(0.079)	(0.076)	(0.065)	(0.019)				
Option physics	5.911	0.467***						
	(449.399)	(0.106)						
Grade in Biology					−0.150*	−0.019*	−0.118*	−0.022*
					(0.079)	(0.010)	(0.065)	(0.012)
Grade in Geology					0.009	0.001	0.070	0.013
					(0.079)	(0.010)	(0.062)	(0.012)
Option geology					1.930***	0.504***		
					(0.520)	(0.158)		
Constant	0.267		1.885		0.482		−0.462	
	(1.538)		(1.400)		(1.666)		(1.302)	
Observations	56	56	56	56	79	79	79	79
Log Likelihood	-17.630	-17.630	-28.076	-28.076	-19.187	-19.187	-27.584	-27.584
Akaike Inf. Crit.	45.261	45.261	64.153	64.153	48.373	48.373	63.168	63.168

*p<0.1; **p<0.05; ***p<0.01

Note: odd columns are the results of the probit regressions, while even columns are the marginal effects of probit regressions, estimated at average.