



HAL
open science

Essays on behavioral economics of confidence, creativity and education

Marco Antonio Gazel Junior

► **To cite this version:**

Marco Antonio Gazel Junior. Essays on behavioral economics of confidence, creativity and education. Economics and Finance. Université Panthéon-Sorbonne - Paris I, 2017. English. NNT : 2017PA01E030 . tel-01798301

HAL Id: tel-01798301

<https://theses.hal.science/tel-01798301>

Submitted on 23 May 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ PARIS 1 - PANTHÉON SORBONNE
U.F.R DE SCIENCES ECONOMIQUES
CENTRE D'ECONOMIE DE LA SORBONNE

THÈSE

POUR L'OBTENTION DU TITRE DE DOCTEUR EN SCIENCES
ECONOMIQUES

présentée et soutenue publiquement

le 28 juin 2017 par

Marco GAZEL

**ESSAYS ON BEHAVIORAL ECONOMICS OF CONFIDENCE,
CREATIVITY AND EDUCATION**

Directeur de thèse

Louis LÉVY-GARBOUA Professeur à l'Université Paris 1 Panthéon-Sorbonne

Jury

Rapporteurs : Claude MONTMARQUETTE Professeur à l'Université de Montréal
Wim GROOT Professeur à l'Université de Maastricht

Président : Jean-Christophe VERGNAUD Directeur de Recherche CNRS - CES

Examineurs : Todd LUBART Professeur à l'Université Paris-Descartes
Giuseppe ATTANASI Professeur à l'Université de Lille 1

L'Université Paris I Panthéon-Sorbonne n'entend donner aucune approbation, ni improbation aux opinions émises dans cette thèse; elles doivent être considérées comme propres à leur auteur.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisor and co-author Prof. Louis Lévy-Garboua for the continuous support of my Ph.D study and related research, for his generosity, patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Claude Montmarquette, Prof. Wim Groot, Dr. Jean-Christoph Vergnaud, Prof. Todd Lubart and Prof. Giuseppe Attanasi, for their insightful comments and encouragement, but also for the hard question which helped me to widen my research from various perspectives.

I acknowledge also the whole structure of the *Centre d'Économie de la Sorbonne*, Paris School of Economics and Université Paris 1 - Pantheon-Sorbonne, which strongly supported me during this long journey. I'm grateful to professors and researchers with whom I had the opportunity to discuss my work and to have exchanges that were fundamental in my development as a researcher: Fabrice Le Lec, Nicolas Jacquemet, Jean-Christoph Vergnaud (again), Stéphane Gautier, Hélène Huber, David Margolis, among many others. In addition, the administrative staff, who helped make this challenge easier and lighter: Loïc Sorel, Audrey Bertinet, Annie Garnerio and Francesca di Legge. My sincere thanks also goes to those who helped implementing my experiments: Maxim Frolov for programming them, Gabrielle Tallon and Yves-Paul Auffrey for helping me to conduct them, Muniza Askari for sharing her work with me, and Marie Thillot for giving credibility to my creativity scores.

In my daily work I have been blessed with a friendly and cheerful group of fellow students. Firstly I would like to acknowledge my friend and co-author Anna Bernard, almost my second advisor. Thank you very much for our crowdfunding projects, discussions and articles. I extend my gratitude to all other (ex-) Ph.D students, with whom I shared my days in the last years: Remi Yin, Noemi Berlin, Leontine Goldzhal, Matthieu Cassou, Antoine Hémon, Antoine Prevet, Antoine Malezieux, Justine Jouxte, Marine Hainguerlot, Quentin Couanau, and Feryel Bentayeb.

The most important learning I had with my thesis is the importance of the family

environment in the development of individuals. I could not have better luck. I won the lottery of birth. I offer my sincerest gratitude to my father, my mother and my sisters. Thanks a lot for the unrestricted support I have always had in my life. Thank you for supporting me despite all the distance. I undoubtedly owe an immense portion of this work to you. Obrigado.

Finally, perhaps the most important acknowledgment. I want to thank Claudia for her complicity and understanding during this entire thesis period. Thank you for making my life more joyful. *Merci beaucoup mon amour.*

À ma petite soeur

Avertissement

Mis à part l'introduction et la conclusion de cette thèse, les différents chapitres sont issus d'articles de recherche rédigés en anglais et dont la structure est autonome. Par conséquent, des termes "papier" ou "article" y font référence, et certaines informations, notamment la littérature, sont répétées d'un chapitre à l'autre.

Notice

Except the general introduction and conclusion, all chapters of this thesis are self-containing research articles. Consequently, terms "paper" or "article" are frequently used. Moreover, some explanations, like corresponding literature, are repeated in different places of the thesis.

Table of Contents

General Introduction	1
1 The hazardous definition of noncognitive skills	1
2 Which <i>noncognitive</i> skills are the most important for education, labor market and life outcomes?	3
2.1 Self-confidence	5
2.2 Creative potential	8
2.3 Motivation and effort	12
3 How school systems impact educational decisions, educational outcomes and intergenerational mobility?	14
3.1 What is an efficient school system?	14
3.2 How to reduce the social gap in educational achievement?	19
4 On the use of experimental methods to study noncognitive abilities and educational institutions	21
5 Outline of the dissertation	23
I Self-confidence and Creativity	29
1 Confidence biases and learning among intuitive Bayesians	31
1 Introduction	32
2 The experiment	36
2.1 Task and treatments	36
2.2 Experimental sessions	38
2.3 Descriptive statistics	39
2.4 Confidence judgments	41
3 Describing confidence biases and learning	42

3.1	Limited Discrimination	42
3.2	Miscalibration	44
3.3	The ability effect	46
3.4	Learning	48
4	Theory	52
5	Predicting confidence biases and learning	57
5.1	Confidence updating by intuitive Bayesians	57
5.2	Regression analysis	61
5.3	Why do intuitive Bayesians make wrong (and costly) predictions of performance?	63
6	Conclusion	66
A	Appendices	69
2	Creative Potential and economic behavior	71
1	Introduction	72
1.1	Creativity in economic literature	72
1.2	An economic approach to creativity	74
2	The experiment	77
2.1	Task and treatments	77
2.2	Descriptive statistics on creativity scores	82
3	Testing economic predictions	85
3.1	Typing: a simple repetitive task	86
3.2	Buttons: a creative exploratory task	93
3.3	Satisfaction	96
4	Creativity and school achievements	99
5	Conclusion	103
A	Appendices	105
A.1	Tables and figures	105

A.2	Holland's typology	114
A.3	Buttons task	115
A.4	Big Five (TIPI)	116
A.5	Creativity scores	117
A.6	Buttons task	120
A.7	Typing task	121
II	School systems	123
	School systems: introduction and experimental design	125
1	A minimal school system	126
2	Methodology	127
2.1	Experimental Design	128
2.2	Experimental sessions	133
3	Descriptive Statistics	133
A	Appendices to Part II	135
A.1	Race treatment	135
A.2	Instructions in French	137
3	An experimental comparison of the efficiency of school systems	143
1	Introduction	144
2	Experimental design, experimental sessions and descriptive statistics	147
3	Efficiency of school systems	147
3.1	Performance in the task	147
3.2	The effect of selection by ability	149
3.3	The effect of self-selection of educational track	149
4	Why choice and early competition fail	152

4.1	Why choice of education fails: maximizing expected utility or minimizing regret?	153
4.2	Why early competition fails: winning in the short run or in the long run?	155
5	Confidence, motivation and performance	157
5.1	Definitions	157
5.2	Confidence	159
5.3	Performance and effort	161
5.4	The choice of track	162
6	The effect of financial incentives	164
7	Testing theoretical predictions	166
7.1	Global performance	166
7.2	Compulsory level	167
7.3	Post-compulsory levels	168
7.4	Self-selection of track	170
8	Conclusion	171
A	Appendices	175
A.1	Tables and figures	175
A.2	Decisions after level 1	179
A.3	Performance and effort (complete version)	181
4	Sorting and socioeconomic bias of school systems	185
1	Introduction	186
2	Experimental design, experimental sessions and descriptive statistics . .	188
3	On the interaction of ability with sorting mechanisms and how it affects the socioeconomic bias of school systems	189
3.1	Ability level	189
3.2	How do school sorting systems interact with ability?	191

3.3	Which school system is more socioeconomically biased?	193
4	The socioeconomic-biasing effect of choice explained	196
5	Gender differences on educational achievement	200
6	The effect of financial incentives	202
7	Conclusion	206
A	Appendices	212
A.1	Tables and figures	212
A.2	Model of rational decisions	215
General Conclusion		221
1	Noncognitive abilities are important for life success	221
2	On the use of experimental economics to investigate school systems	223
3	Understanding the efficiency of school systems and inequalities on educational attainment	225
4	Education policy agenda must address noncognitive abilities	227
List of Tables		249
List of Figures		252
Résumé substantiel		255
1	La définition hasardeuse des compétences non cognitives	255
2	Quelles compétences non cognitives sont les plus importantes pour l'éducation, le marché du travail et les réussites de la vie?	257
2.1	La confiance en soi	260
2.2	Le potentiel créatif	263
2.3	Motivation et effort	267
3	Comment les systèmes scolaires influencent les décisions éducatives, les résultats scolaires et la mobilité intergénérationnelle?	269

3.1	Qu'est-ce qu'un système scolaire efficace?	270
3.2	Comment réduire l'écart social dans les résultats scolaires? . . .	275
4	L'utilisation de méthodes expérimentales pour étudier les capacités non cognitives et les établissements scolaires	277
5	Aperçu de la thèse	280
	Résumés	285

GENERAL INTRODUCTION

"Some kids win the lottery at birth; far too many don't - and most people have a hard time catching up over the rest of their lives. Children raised in disadvantaged environments are not only much less likely to succeed in school or in society, but they are also much less likely to be healthy adults."

James Heckman

This thesis contributes to the growing economic literature on *noncognitive* skills that are critical for life success, specially for academic success. It comprises four essays based on behavioral and experimental economics approaches, with two main objectives. The first objective is to study two noncognitive skills, namely self-confidence and creativity. We aim at understanding the determinants of self-confidence, and the impact of creative potential on economic outcomes. The second objective is to study how school systems impact educational decisions, educational outcomes and intergenerational mobility, where noncognitive skills may play an important role, specially self-confidence and motivation. This introduction puts in perspective the questions and concepts developed within each chapter and provides an overview of the thesis.

1. The hazardous definition of noncognitive skills

The term noncognitive did emerge in the economic literature in the early 2000s with the interest in explaining the variability on educational, labor market and other economic outcomes that was unexplained by measures of cognitive skills. Cognitive skills are measured by intelligence tests, school grades or standardized tests (see [Bowles](#),

Gintis and Osborne, 2001, Heckman and Rubinstein, 2001, for instance), which measure intelligence and knowledge. However, the identification, classification and measurement of noncognitive skills is still a challenge for economists ¹ (Humphries and Kosse, 2017). Thus, insights from other sciences may bring important cues for economists.

Neuroscientists explain that most of these skills depend on the executive functions of the brain (Diamond and Lee, 2011). The later refers to a family of mental functions (cognitive control) that are needed when the agent has to concentrate and think instead of going "on automatic", relying mainly on prefrontal cortex. The three core executive functions are inhibitory control ², working memory ³, and cognitive flexibility ⁴ (Diamond, 2013). High-order executive functions are problem solving, reasoning and planning, which are related to fluid intelligence ⁵. Thus, it seems obvious that the skills, so called as noncognitive by economists, depends also on the cognition ⁶.

In our opinion, the distinction between cognitive and noncognitive skills is a misnomer because it is based on a false dichotomy. There are no measures of cognitive skills that don't at least partly reflect noncognitive factors of motivation and context, while measures of noncognitive skills will likewise be dependent on cognitive and situational factors. So talking about noncognitive skills can be misleading. Even though, this thesis doesn't aim at bringing a new terminology for these skills. We keep the terminology noncognitive, as we can relate to the literature. We however provide a precise definition: noncognitive skills correspond to abilities that are important for life success but are different from knowledge (measured by achievement tests) and IQ.

1. The extensive list of terminologies for noncognitive skills found in the economic literature illustrates the difficulty to define them. This literature includes such terms as noncognitive abilities, soft skills, socio-emotional skills, behavioral skills, character, and personality traits.

2. Which includes self-control, discipline and selective attention.

3. Holding information in mind and manipulating it, which is essential for reasoning.

4. Including creative problem solving and flexibility.

5. Fluid intelligence is the ability to deal with novel problems, independent of any knowledge from the past. It is considered one of the most important factors in learning (Jaeggi et al., 2008).

6. The American Psychological Association Dictionary defines cognition as "all forms of knowing and awareness such as perceiving, conceiving, remembering, reasoning, judging, imagining, and problem solving."

However, noncognitive abilities are impacted by them. These abilities vary according to situation, and importantly, can be improved during life span. We thus consider self-confidence, creativity and motivation as noncognitive skills. These skills are discussed in the next section.

2. Which *noncognitive* skills are the most important for education, labor market and life outcomes?

What matters for life success? A growing literature bringing insights from psychology and sociology to economic theory shows that education, labor market and life outcomes depend on many skills, not just the cognitive skills measured by IQ, grades, and standardized achievements tests (Borghans et al., 2008). Heckman and Kautz (2012) show that measures of cognitive skills during adolescence explain less than 15% of hourly wage at age 35. It suggests that grades are not only determined by hard skills. For instance, evidence shows that discipline accounts for over twice as much variation in final grades as does IQ, even in college (Duckworth and Seligman, 2005). Thus, the study of noncognitive skills became an important topic for economists in the past fifteen years, since Bowles, Gintis and Osborne's (2001) seminal survey of the determinants of earnings -a milestone in the economic literature of noncognitive skills. The extensive list of noncognitive skills that are important for economic outcomes includes self-confidence, respect for others, ability to build consensus, willingness to tolerate alternative, academic motivation, academic confidence, persistence, communication skills, creativity, and teamwork, among many others (Heckman, 2011, García, 2016). Given the relative novelty of the field for economists, this list is likely to grow as more evidence emerges.

The direct effects of noncognitive skills are important for every aspect of life – success in school and in the workplace, marital harmony, and avoiding things like smoking,

substance abuse, or participation in illegal activities (Heckman, Stixrud and Urzua, 2006). The indirect effects of these skills also matter, mainly because noncognitive skills favor cognitive developments. In other words, the development of noncognitive abilities in effect improve academic skills such as reading, writing, and mathematics performance (García, 2016).

The development of noncognitive skills starts in early infancy and has important influence from family and societal characteristics (Cunha, Heckman and Schennach, 2010, Diamond, 2013, García, 2016). Through socialization, educated parents automatically transmit their capabilities and preferences to their children (Bourdieu and Passeron, 1964, Becker and Tomes, 1979). For instance, parentally supplied verbal environment of children at age of three years old strongly predicts reading comprehension at 10 years of age (Hart and Risley, 1995). Additionally, children from upper social classes benefit from the environment they grow in since they have later contact with violence, death, drugs and criminal justice system, and a positive precocity in recognizing letter and numbers, knowing other neighborhoods and cities, and reading newspaper headlines when compared to children from lower social classes (Farah, Noble and Hurt, 2006). Noncognitive abilities are however unequally distributed, as a mirror of social inequalities. Assuming equally distributed innate abilities, noncognitive skills inequalities may explain the persistence of social inequalities⁷.

Given that executive functions -and consequently noncognitive skills- can be improved through life span, early training might be an excellent mean to reduce inequality. Indeed, there is scientific evidence supporting the improvement of executive functions (and noncognitive skills) in the early school years. For instance, Heckman, Pinto and Savelyev (2013) show the positive effect of the Perry Preschool Program on the development of noncognitive skills of low-income children⁸. Diamond and Lee (2011) show

7. Assuming that difference in opportunities can be neutralized by public policies.

8. The Perry Preschool Program (1962-1967) provided high-quality preschool education to three years old low income African-American. The program did not produce gains in the IQ of participants (modest gains for women, no gains for men), but improved scores on achievement tests at age 10. This

that interventions in the early school like computerized training, aerobic exercise and martial arts enhance the development of executive functions with a special benefit for less-advantaged children⁹, reinforcing the potential of early education in heading off gaps in achievement between more- and less-advantaged children.

Current education policy focuses on cognitive skills. Less room is left to improve noncognitive skills, even though they can be improved in schools (Blair and Razza, 2007), leading to an increased intergenerational mobility. The literature on noncognitive skills (and executive functions) suggests that education has multiple dimensions, encompassing skills and attitudes, not just intelligence and knowledge (Diamond, 2013). In “*The Need to Address Non-Cognitive Skills in the Education Policy Agenda.*”, García summarizes noncognitive skills that schools should develop and policies should promote:

"[...] these include critical thinking skills, problem solving skills, emotional health, social skills, work ethic, and community responsibility. Also important are factors affecting personal relationships between students and teachers (closeness, affection, and open communication), self-control, self-regulation, persistence, academic confidence, teamwork, organizational skills, creativity, and communication skills".

Next subsections put in perspective the economic importance, advances in the research and measurement of the noncognitive skills addressed in this thesis: self-confidence, creative potential and motivation.

2.1. Self-confidence

Subjective beliefs are important for all situations where an economic agent makes decisions under uncertainty. The decision maker assigns subjective probability estimates for each state of nature involved in the decision, choosing the one that maximum result confirms the importance of (improving) noncognitive skills for academic success.

9. Lower income, lower working-memory span and attention deficit hyperactivity disorder (ADHD) children show the most executive functions improvement from these interventions.

mizes her expected utility. When the true probability is unknown, the agent estimates her subjective probability updating prior information about herself and about the environment, like Bayesians do (Van den Steen, 2011, Möbius et al., 2014). From a standard theoretical perspective, confidence is a distorted probability of success that can be updated according to experience and available information. Indeed, the impact of self-confidence over agents' behavior goes beyond the decision process, because it also impacts the motivation -and so the effort- to perform the task increasing the likelihood to succeed. Self-confidence also acts as an incentive to reinforce and maintain one's self-esteem (Bandura, 1993, Bénabou and Tirole, 2002).

This thesis considers a specific type of confidence: the self-confidence, that are the beliefs an agent holds about his own ability. In many circumstances, people appear to be overconfident in their own abilities whatever the difficulty of task, i.e. their subjective probability of success is higher than the normative chances to succeed the task. Moore and Healy (2008) identify three different forms of overconfidence as overplacement, overestimation, and overprecision. Overplacement¹⁰ occurs when individuals compare themselves with others, massively finding themselves "better-than-average" in familiar domains (eg., Svenson 1981, Kruger 1999). The overestimation is the most common in the literature, it takes place when agents overestimate their own absolute ability to perform a task (eg., Lichtenstein and Fischhoff 1977, Lichtenstein, Fischhoff and Phillips 1982). Finally, the overprecision arises when people overestimate the precision of their estimates and forecasts (eg., Oskamp 1965). This dissertation aims at understanding how people overestimate, or sometimes underestimate, their own absolute ability to perform a task in isolation.

The estimation of self-confidence to perform a specific task depends on cognitive ability and other individual characteristics (Stankov, 1999). Thus, the higher is the cognitive ability, the lower is the estimation bias (Stankov et al., 2012). Gender is

10. Literature also refers to overplacement as "better-than-average" instead of overconfidence.

one example of an individual characteristic that affects overconfidence. [Stankov et al. \(2012\)](#) show that girls present a lower estimation bias than boys even if they report the same level of confidence on mathematics and English achievement tests¹¹. Another individual characteristic that impacts self-confidence is the family background. Using cross national data (PISA), [Filippin and Paccagnella \(2012\)](#) show evidence of the positive relation between the family background and academic confidence reported by 15 year-old pupils. For a given level of ability, the higher is the socio economic status, the higher is the subjective confidence to succeed at school.

If on one hand overestimation may have a negative impact leading individuals to non-optimal decisions, on the other hand it enhances one's effort, increasing then performance and the probability of success. When coupling the effects of confidence on decisions and performance to the impact of family background to the level of self-confidence, [Filippin and Paccagnella \(2012\)](#) asserts that: "self-confidence can be a channel through which education and earning inequalities perpetuate across generations."

We elicit self-confidence using a self-report measure. Individuals are directly asked to state their probability of success for a given task, for instance: "what are your chances of success on the scale of 0 to 100?". The Adams's ([1957](#)) scale is convenient for quantitative analysis because it converts confidence into (almost) continuous subjective probabilities. Self-report methods have been widely used and validated by psychologists and neuroscientists; and recent careful comparisons of this method with the quadratic scoring rule¹² found that it performed as well ([Clark and Friesen, 2009](#)) or better ([Hollard, Massoni and Vergnaud, 2015](#)) than the quadratic scoring rule¹³.

11. In other words, boys perform more poorly than girls even though they are about as confident as girls.

12. After the subject has reported a probability p , the quadratic scoring rule imposes a cost that is proportional to $(1 - p)^2$ in case of success and to $(0 - p)^2$ in case of failure. The score takes the general form: $S = a - b$. Cost, with $a, b > 0$.

13. The second study also included the lottery rule in the comparison and found that the latter slightly outperformed self-report. The lottery rule rests on the following mechanism: after the subject has reported a probability p , a random number q is drawn. If q is smaller than p , the subject is paid

Psychologists have developed several scales to elicit measures of self-confidence for specific domains, which are not used in this thesis. For instance, the Academic Behavioural Confidence scale (Sander and Sanders, 2003, 2006) provides a measure of academic confidence and self-efficacy¹⁴. Note that this scale measures also self-efficacy. Indeed, self-efficacy is considered a good proxy for self-confidence¹⁵, regularly found in the confidence literature (see Stankov et al., 2012, for instance).

2.2. Creative potential

Creativity has been defined as “the ability to produce work that is both novel and appropriate” (Sternberg and Lubart, 1996), which is a substantial drive for innovation. According to Feinstein (2009) "creativity and its counterpart innovation are the root of progress and thus fundamental to the dynamics of economic systems". Indeed, several theories attribute technology innovation to the strong economic growth after the Second World War (Romer, 1986, for instance). Thus, the potential of creativity -i.e. the potential to produce creative works- should be an important topic of interest for economists. However, few economic studies have dealt with creative behaviors so far. The existing economic literature lies in the fact that production and consumption of new products are uncertain activities, implying risk taking and entrepreneurial skills associated with the creative behavior (Menger and Rendall, 2014). More recently, Charness and Grieco (2013) studied the effect of incentives on the production of creative works. But the impact of the potential of creativity on economic outcomes is still a lack in the economic research¹⁶.

according to the task. If q is greater than p , the subject is paid according to a risky bet that provides the same reward with probability q .

14. Another example: the Activities-specific Balance Confidence Scale (Powell and Myers, 1995) used in the medical domain measures the confidence in performing various ambulatory activities without falling or experiencing a sense of unsteadiness.

15. Self-efficacy was defined by Bandura (1986, page 391) as "people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performances".

16. This impact of the potential of creativity on economics outcomes is addressed by this thesis in Chapter 2.

The scarcity of research on creative behavior in the economics field can be attributed to two main factors. First, economists neglected the impact of noncognitive abilities -thus, creativity- on economic outcomes up to the last decade (Borghans et al., 2008). Second, the assessment of creativity potential is still a bottleneck in the economics research, these measures are not yet applicable in large scales, being restricted to experimental studies. The recent interest for noncognitive abilities and the evolution of research on creativity by psychologists in the last three decades¹⁷ however put creativity in the spotlight on research in economics.

Guilford (1950) work was a turning point in the psychological research on creativity. Up to this seminal work, creativity was associated to an exceptional process of gifted individuals. Thus, the assessment of creativity was not an important issue since 'creativity' was directly observable by the production of artists (Barbot, Besançon and Lubart, 2011). In the context of the period post Second World War -which required innovation in research and development-, Guilford (1950) claimed that the potential of creativity is not restricted to gifted individuals, and importantly, can be measured and developed. Creativity would thus be considered as a cognitive and social process, not only a personality trait. Indeed, the creative potential depends also on domains¹⁸ and tasks (Lubart and Guignard, 2004).

Theories developed by psychologists in the last decades confirmed Guilford's proposition: a creative behavior depends on many factors. Sternberg and Lubart (1995) propose a multivariate approach, for which creativity is influenced by cognitive (intelligence and knowledge), conative (motivation, personality traits and thinking style) and environmental factors¹⁹. In the same vein, the investment theory of creativity enu-

17. Barbot, Besançon and Lubart (2011) argue that in "the 90's, the creativity research literature increased exponentially with the appearance of new scientific journals, international conferences and book series on the topic, which coincided on the other hand, with significant progress in psychometric science."

18. Examples of creative domains: graphic-artistic, verbal-literary, social problem solving, musical and creative (Lubart, Zenasni and Barbot, 2013).

19. Emotional factors can be added to this list (Lubart et al., 2003)

merates six distinct but interrelated resources required to enhance creativity, namely intellectual abilities, knowledge²⁰, styles of thinking, personality, motivation, and environment (Sternberg and Lubart, 1991*b*, Lubart and Sternberg, 1995, Sternberg, 2006). The neurological perspective attributes to the frontal and the prefrontal cortex, thus to executive functions, the central role in the creative process (Borst, Dubois and Lubart, 2006).

These theories show mainly that the potential of creativity can be enhanced. Consequently, schools have an important influence on the development of creative behavior. The learning environment and pedagogy have a direct impact on the development of the creative potential, which effect is higher for children with lower initial creative potential (Besançon and Lubart, 2008). For instance, Sternberg and Lubart (1991*a*) show that alternative pedagogies, like Montessori and Freinet, can develop divergent thinking, an important component of creativity. In this context, authors assert that "schooling can create creative minds - though it often doesn't". Thus, developing a learning environment to enhance creativity, considered by the National Research Council (2013) as one of the key skills necessary for 21st century learning outcomes, seems an important goal of education.

In psychology, the creative decision process is decomposed into a phase of mental divergence followed by a phase of mental convergence. Mental divergence allows finding new ideas to problems; while mental convergence allows the synthesis of disparate ideas into a novel and appropriate solution. Obviously, both traits are useful for innovating and must act in coordination because new ideas don't fall from heaven, they come to the mind. Many things come to the mind all the time though, but, if the person is focused in a specific direction, she might lend attention to a signal and convert it into a valuable idea if she is endowed with a good sense of serendipity. What seemed to indicate mental

20. Knowledge can both promote or inhibit creativity. On one hand it is impossible to innovate in a field if one has no knowledge about it, on the other hand a lot of knowledge about a field can result in a closed perspective about it (Sternberg, 2006).

divergence, i.e. finding new ideas, requires a form of mental convergence. Things that come to the mind are not automatically interpreted as signals: most of them will be probably dismissed as noise and forgotten, and only appropriate target-directed ideas will be recalled. Thus, divergence and convergence are complementary within the creative personality.

The assessment of creative potential is still a big challenge to integrate this variable into economic models and theories for two reasons. First, it takes time to measure creativity, and measurements better be done within the lab, or in any other controlled environment, such as a classroom. The most reliable and complete measures use the production-based approach, in which individuals are asked to produce a work in a given creative domain. A comparison with the production of other individuals provides a measure of creativity (see for instance [Charness and Grieco, 2013](#)). Second, the assessment depends on the domain of creative productions (graphical or verbal) and modes (divergent or convergent thinking). A reliable and complete example of creative potential assessment among children at school is the [Lubart, Besançon and Barbot's](#) EPoC battery (*Evaluation du Potentiel Créatif*, 2011), which measures the potential of divergent and convergent thinking in two different domains, namely graphic-artistic and verbal-literary. This procedure, which is used in this thesis, has a great validity: authors found a high and significant correlation between divergent thinking and the traditional Torrance's test of creative thinking ([Torrance, 1962](#)). Moreover, they found a correlation between the creativity measures of the EPoC to *openness* personality trait, in line to [McCrae and Costa \(1987\)](#) observations that *openness for new experiences* facilitates divergent thinking. The complete battery of EPoC's test takes around two hours.

2.3. Motivation and effort

Motivation corresponds to the set of internal and external factors that stimulate agents to make an effort to attain a goal. Thus, motivation has an important impact on behavior, including decisions, performance and outcomes. Motivation may explain why agents with different abilities for a given task reach the same outcome, or the other way around, why individuals with identical abilities have different outcomes.

Literature distinguishes between two types of motivation, namely intrinsic and extrinsic motivation. Intrinsic motivation is driven by a personal interest or enjoyment in the task itself, typically associated to high-quality learning and creativity. Extrinsic motivation comes from external influences in order to attain a desired outcome, normally characterized by rewards and penalties. Both, intrinsic and extrinsic motivations, are related to performance, satisfaction, trust, and well-being (Gagné and Deci, 2005). However, the effectiveness of extrinsic motivation to promote a sustainable effort is controversial because in some circumstances it pushes agents to act with resistance and disinterest to achieve imposed goals (Deci and Ryan, 1985, 2010). The impact of reward on intrinsic motivation is also controversial. For a long time the consensus in the social psychology research pointed to a negative impact of rewards on the intrinsic motivation, and so creativity. Alternatively, recent studies show that in some circumstances rewards enhance extrinsic motivation without deteriorating the intrinsic motivation (Gagné and Deci, 2005, Charness and Grieco, 2013), equivalently Hennessey and Amabile (2010) state that "the expectation of reward can sometimes increase levels of extrinsic motivation without having any negative impact on intrinsic motivation or performance". Given the ambiguous effect of extrinsic reward over intrinsic motivation, the Self-Determination Theory (Deci and Ryan, 1985, 2010) distinguishes diverse types of extrinsic motivations, some of which do represent barren forms of motivation and some of which represent active forms of motivation (Ryan and Deci, 2000).

The concepts of autonomous and controlled motivations arise from this distinction.

Controlled motivation is driven by experiences of pressure and obligation, limiting the desired behavior to the period when the external regulation is present. On the contrary, autonomous motivation, also known as sustainable motivation, satisfies the human needs for competence, relatedness and autonomy. It is consistent with intrinsic motivation, providing to individuals the sense of choice, volition, and self-determination (Stone, Deci and Ryan, 2009). Thus, autonomous motivation may have an important impact on educational outcomes. Gagné and Deci (2005) suggest that:

"[...] because many of the tasks that educators want their students to perform are not inherently interesting or enjoyable, knowing how to promote more active and volitional (versus passive and controlling) forms of extrinsic motivation becomes an essential strategy for successful teaching".

We measure motivation by effort, that is the amount of time used to perform a real effort task in the lab, i.e. solving anagrams. Motivational variables are related to effort by definition (Brookhart, Walsh and Zientarski, 2006), thus a measure of effort is the best proxy to measure motivation. However, it is always complex to have a precise measure of effort at school or at the workplace (Taylor and Taylor, 2011). In order to close this gap, psychologists have developed several self-reported based scales to assess motivation when it is not possible to have a precise measure of effort. These scales are based on other variables linked to the concept of motivation, such as self-esteem, self-efficacy, self-regulation, locus of control and goal orientation. For instance, the Academic Motivation Scale (Vallerand et al., 1992) is developed to measure intrinsic and extrinsic motivation in education.

This section has presented the importance of noncognitive skills for economic outcomes, and has provided a literature review of the skills addressed by this thesis. Importantly, these skills can be developed during the life cycle, however the early development has an important impact of the family background characteristics. Thus, according to a strong body of research, these skills may explain (at least partially)

the persistence of educational inequalities. The next section discusses the efficiency of school systems, another potential source of socioeconomic biases - which is addressed in Part II of this thesis.

3. How school systems impact educational decisions, educational outcomes and intergenerational mobility?

Tests in maths, science and reading in the Program for International Student Assessment (PISA) show that students' average achievement level varies considerably across countries. [Wößmann \(2016\)](#) argues that different school systems are responsible for a considerable portion of the cross-country achievement variation. Each country has its own school system, which comprises a set of educational institutions that are shaped by public policies. The cross-country comparison shows that educational institutions such as tracking and ranking have an important impact on pupils' decisions and outcomes. For instance, [Wößmann \(2016\)](#) shows that early tracking into different school types by ability increases educational inequalities, without increasing achievement levels²¹. The question is thus to identify which are the most efficient educational institutions.

3.1. What is an efficient school system?

The concept of efficiency is quite cloudy for school systems. In this thesis, school systems' efficiency implies an economic state in which every resource is optimally allocated, serving each agent in the best way. In other words, an efficient school system maximizes students' educational expected outcomes. Equivalently, an efficient school

21. An experimental comparison of the efficiency of different school systems is one of the objectives of this thesis. It is presented in [Chapter 3](#)

system should minimize expected regret²², which is associated to failures and dropouts. Students who failed and suffered an opportunity loss will regret their choice ex-post and ask for political redistribution. While political platforms aim at reducing failures -viewed as wastage, or ex-post inefficiency-, students pursue their own objective of maximizing expected utility (EU) which, unfortunately, does not guarantee future success and may cause regret.

What guides educational decisions? Choosing is not a psychologically simple task. Important decisions such as educational decisions²³, or even more trivial decisions, can lead to regret and concern over missed opportunities causing dissatisfaction even with good decisions (Schwartz, 2004). Therefore, pupils must learn how to choose so as to minimize regret.

Educational decisions follow the expected utility hypotheses. Agents are supposed to weight the prospects of a given decision with their probability of success, and select the alternative with the highest expected utility. Equivalently, Heckman, Humphries and Veramendi (2016) assert that:

"[...] in modern parlance, individuals should continue their schooling as long as their ex-ante marginal return exceeds their ex-ante marginal opportunity cost of funds".

Here, it is important to posit that even if a given educational decision is efficient ex-ante because it maximizes the ex-ante expected utility, it can be inefficient ex-post causing regret since there is an increasing risk of failure for higher levels of education.

22. In expected utility (EU) theory, it is known that for all A, B : $EU(A) - EU(B) = EOL(B) - EOL(A)$, with $EOL(A)$ designating the expected opportunity loss of A with respect to B (see Raiffa, 1968, for instance). If $EOL(A)$ is the measure of expected regret of choosing A and foregoing B . So, maximize EU is equivalent to minimize expected regret. Although the two programs are equivalent by duality, it is more common to speak of EU maximization. In the context of education policies, the value of speaking of regret is because ex-post regret feeds relative frustrations and political discontent. However, the two objectives: maximize EU and minimize expected regret yield the same conclusions.

23. Pupils must decide for the extension of their education, i.e. decide to start a new schooling year or to go to the job market. Students decide also for the track, e.g. general (or academic), vocational, or technical track. The variety of tracks and the timing of these decisions vary across school systems.

In modern societies, educational decisions are probably the most important choices faced by individuals during their life cycle. They are important because schooling has a strong influence in the life-span monetary and non monetary outcomes (Heckman, Humphries and Veramendi, 2016). The positive effect of education goes beyond market outcomes impacting also future behavior, such as health behavior, smoking, drugs consumption, fertility, household management, savings, among others (see Vila, 2000, Lance, 2011, Król, Dziechciarz-Duda et al., 2013, Heckman, Humphries and Veramendi, 2016, among others.).

Educational choices are probably the hardest decisions too, thus education is the life domain with highest potential of regret in contemporary society (Roese and Summerville, 2005). They are hard for two main reasons. First, as discussed in the previous paragraph, education has important monetary and non monetary consequences in one's future. Therefore, estimating returns to schooling when deciding is a complex challenge for agents. Second, because these decisions are surrounded by great uncertainty, since pupils have an imperfect knowledge about their ability and preferences when deciding. Consequently, family and social environment play an important role and influence in educational choices.

What predicts the normative probability of success at school? In the educational context, the prospect of high wages for higher education may push pupils to rationally try higher levels of education, even those with low chances of success. However, rational agents take into account their probability of success before making their decision in order to avoid failure and regret, so a good estimate of future chances of success is crucial for optimal decisions. The question then arises as to *what predicts success at school?* Recent advances in the economic literature claim that educational achievement depends on an extensive set of cognitive and noncognitive skills, such as: motivation persistence, self-esteem, risk tolerance, optimism and time preferences (see Heckman and Rubinstein, 2001). For instance, Castillo et al. (2011) found that more

patient children, those who are concerned with future consequences of their behavior, have a more favorable outlook.

There is no disagreement that intelligence, as measured by IQ, is an important predictor of success at school. Higher IQ increases the performance at school, but it is not a guarantee of future success if the pupil is not motivated by her studies. Achievement motivation is also important for academic success (Busato et al., 2000) and has a direct impact on the students' perception that success depends on one's effort (Ames and Archer, 1988). Effort (and motivation) may explain why students with different cognitive ability levels can reach the same educational outcome.

In the same vein, we cannot neglect that actual grades are shaped by cognitive and noncognitive abilities. Academic record may give important signals for pupils about their future performance. The problem of basing decision on academic record is that this measure does not account for the increasing level of difficulty in education. Moreover, it does not guarantee future motivation if pupils do not decide according to their preferences, that are under development. According to Schwartz (2009), a pupil with doubt about whether she made the right educational choice, may be less engaged to her studies than someone who lacks such doubts. Less effort is likely to translate into worse performance.

Summing up, even if there are important cues to predict future academic success, it is unlikely that pupils can properly estimate the normative chances of success when they make their educational choices. Here, it is important to postulate that since pupils don't know ex-ante their true probability of success for further studies, educational decisions are based on their subjective probability to succeed, i.e. their self-confidence.

Given the complexity to understand the chances of future educational achievement, public policies may play an important role in helping pupils to match their abilities to optimal decisions, increasing welfare and decreasing regret. Filippin and Paccagnella (2012) bring an interesting example of this kind of educational policy when comparing

Netherlands' and Italy's school systems. The main difference between these two countries concerns the self-selection for the high school tracks²⁴. In Netherlands, results of a nationwide aptitude test at age 12 act as a reference of the most suitable track for pupils aptitudes. Differently, pupils -and their parents- have no signal to select their preferred track in Italy. Figure II.1 shows the effectiveness of signaling students' ability in Netherlands. The lower degree of overlapping across trackings in Netherlands suggests a better matching between ability and educational track when pupils have better signals about their abilities.

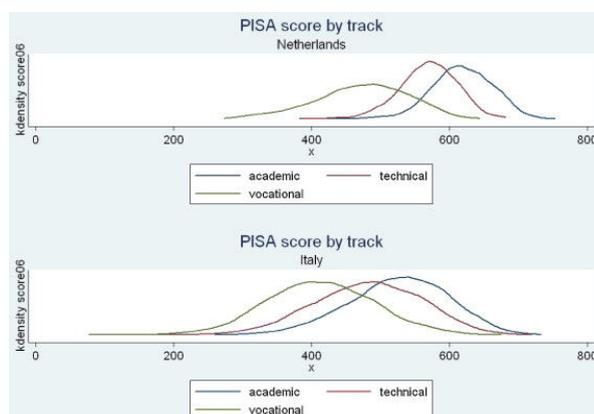


Figure .1 – High school tracking by ability. Reprinted from [Filippin and Paccagnella \(2012\)](#) with permission.

In the same vein, [Goux, Gurgand and Maurin \(2016\)](#) show a randomized controlled trial in France, in which low-achievement students and their families have had several meetings with school principals during middle school. The aim of these meetings was to explain: (i) the importance of choices they should do by the end of the academic year²⁵, and (ii) that the actual performance of pupils should be more important for educational decisions than family aspirations. This program helped pupils (and families) to formulate educational objectives better suited to their academic aptitudes, shaping

24. The three possible tracks here are academic (general), vocational and technical.

25. In France, students must decide for the high school track (vocational or academic) by the end of middle school

the high school decisions of the less realistic students. Consequently, this program reduced failures at high school by 25% in the target population.

This section shows that besides having greater chances of success, more able students are likely to make more suitable educational choices. These two observations together may be one of the causes of inequality observed in school success, which is discussed in the next section.

3.2. How to reduce the social gap in educational achievement?

The importance of education for intergenerational mobility is a consensus for social scientists (for example [Piketty, 2000](#), [Black and Devereux, 2011](#)). Thus, understanding the causes of inequalities in educational achievement is an important issue to promote intergenerational mobility.

Socioeconomic bias We define socioeconomic bias in education by the degree to which educational decisions and educational attainment is impacted by pupils' socioeconomic status. Under this assumption, the more favorable to upper ability groups is a school system, the more socioeconomic biased the system is. Thus, if children of upper classes in society are overrepresented in upper levels of education, this is an evidence of socioeconomic bias.

The classical analysis of [Becker \(1967\)](#) on intergenerational mobility attributes the inequality of chances essentially to differences in abilities and opportunities. Considering that innate abilities are equally distributed in all social classes, the socioeconomic bias of education vanishes once social differences in opportunities can be neutralized. Human capital theory ([Becker, 1964](#)) demonstrates that an efficient credit market on education investments is all that we need to reach this goal. This optimistic prediction has not quite materialized in developed countries, however, in spite of sustained efforts to eradicate differences in opportunities. Several studies have shown that differences in opportunities played only a marginal role in developed countries ([Carneiro](#)

and Heckman, 2002, Cameron and Taber, 2004).

Therefore, the persistence of inequalities in educational attainment demonstrates the presence of socioeconomic-biased noncognitive abilities, built up during childhood and adolescence. Those noncognitive abilities are inherited by children and youth from their permanent exposition to their parents, friends, peers, and social environment and the differential investment of families in their human capital (Lévy-Garboua, 1973, Becker and Tomes, 1979, Cunha and Heckman, 2008, Heckman and Farah, 2009, Cunha, Heckman and Schennach, 2010). No doubt that upper class children are likely to grow up in better learning environments, with more stimuli²⁶ and less stress (Heckman, 2011). Additionally, children of different socioeconomic status do not have the same educational choices as they do not have the same reference points and aspiration levels: children from lower-SES can consider a success what people from higher-SES consider a failure (Boudon, 1973). James Heckman uses the term "*the Lottery of birth*" when describing the powerful effect of family's legacy in shaping the trajectory of one's life. Indeed, several studies claim that the family background characteristics are more important than school social composition and school resources to predict educational outcomes (Chudgar and Luschei, 2009, Borman and Dowling, 2010).

Considering the development of noncognitive skills in early education can be an equalizing factor in the competition for a selective social position. Schools need to target the early development of noncognitive abilities in order to reduce the effect of family background in the intergenerational mobility. An equitable system should improve the outcomes of less able individuals, without prejudice for the more able.

We have seen that noncognitive abilities are important for life success and may have important impact on educational decisions and performance, being a possible explanation for the persistence of inequalities on educational attainment among classes. This thesis aims at studying three noncognitive abilities, namely creativity, self-confidence

26. For instance, the number of books at home are the most important predictor of academic achievement (Fuchs and Woessmann, 2004).

and motivation and the impact of the two latter on educational decisions and educational inequalities. The next section presents the experimental methods and shows why it is well suited for this thesis approach.

4. On the use of experimental methods to study noncognitive abilities and educational institutions

The measurement of noncognitive abilities is a challenge for economists (Humphries and Kosse, 2017). The assessment of most psychological variables relies in tests and surveys that require time and, preferably, a controlled environment. Thus, the use of lab experiments seems the most suitable alternative to introduce psychological variables in the economic research²⁷. The lab experiment allows us to measure variables that are in the core interest of this thesis, like creativity, performance, ability, self-confidence and effort that would be difficult to observe precisely in surveys.

The main advantage of laboratory experiments is the possibility to isolate specific variables of interest while controlling for the environment. This mechanism allows for the isolation and identification of causal effects. In this context, experimental methods is a powerful tool to test theories, search for new facts, compare institutions and environments, and test public policies.

In general, environments created in the lab are simpler than those found in nature. The question of the external validity of lab experiments then arises, i.e. to what extent in-lab behavior is correlated to real life behavior and the results of a study can be generalized? This issue is controversial among economists. Levitt and List (2007) summarize the main criticisms about the external validity of in-lab measures. Authors argue that in-lab: (i) the context, choice sets and time horizons cannot be completely replicated in-lab, (ii) characteristics of experimental subjects differ from

27. In fact, experimental economics in the lab was inspired by social psychology experiments.

groups engaged in out-lab decisions, *(iii)* monetary incentives are different from real-life. In a critical reply to [Levitt and List \(2007\)](#), [Camerer \(2015\)](#) provides a more favorable outlook for experimental, he argues that: *(i)* external validity is not a primary concern in a typical experiment, since experimental economics aims at establishing a general theory linking economic factors such as incentives, rules and norms to decisions and behavior, *(ii)* some experiments have features that cannot be generalized to the field, like some field settings that cannot be generalized to other field settings, *(iii)* most economic experiments reviewed by him, and summarized in this article, show a correlation between in-lab and out-lab behavior. In the same vein, [Plott \(1991\)](#) argues that in the lab:

"[...] real people motivated by real money make real decisions, real mistakes and suffer real frustrations and delights because of their real talents and real limitations."

Thus, even if it is not possible to replicate natural environments, we can have valuable cues from behaviors and decisions through incentivized lab experiments.

The four chapters of this thesis present incentivized laboratory experiments. Chapters [1](#) and [2](#) are based on real-effort tasks where decision-making and behavior are observed and analyzed in order to study self-confidence and creativity. Chapters [3](#) and [4](#) proposes an experimental stylized educational system that allows the comparison of different school systems and educational institutions.

There are several empirical limitations to study a given institutional context and/or make international comparisons using field data as it is almost impossible to isolate the investigated effect maintaining everything else constant. The use of an experimental framework is a good alternative to study educational institutions, even if it is not possible to capture all the elements of an educational system under a controlled environment. To address this research objective, we reproduce by means of an incentivized

lab experiment the stylized educational system²⁸, where we can discriminate differences in curricula, differences in payoffs, the choice of the track and the performance level to qualify for a certain curriculum.

Experimental results allow an easy and valid comparison of the overall performance of a minimal school system under various sorting mechanisms, thus facilitating the identification of the efficient design that is, the educational output-maximizing design conditional on the ability distribution.

5. Outline of the dissertation

This thesis presents two main parts, respectively organized in two chapters. The first part is dedicated to study two noncognitive skills implied in the decision process, that are the confidence on future success and the potential of creativity. Self-confidence on future success is a topic of interest for economists for a long time since it has been an important determinant of the decision process: individuals are assumed to maximize their expected utility according to their (subjective) probability of the different outcomes they will face. If on one hand self-confidence is widely studied by economists, on the other hand the potential of creativity is a new variable of interest in this research field. The psychological measure of the potential of creativity assesses the extent to which an individual is able to engage in creative work (Lubart, Zenasni and Barbot, 2013), and thus may have an important impact on economic outcomes. The second part of this dissertation focuses on the experimental investigation of school systems, with two main objectives: the comparison of the efficiency of different school sorting mechanisms and an evaluation of social and gender biases caused by them.

Chapter 1 compares the speed of learning one's specific ability in a double-or-quits game with the speed of rising confidence as the task gets increasingly difficult. We find that people on average learn to be overconfident faster than they learn their true ability

28. The stylized educational system used in this research is described in page 127.

and we present an Intuitive-Bayesian model of confidence which integrates these facts. Uncertainty about one's true ability to perform a task in isolation can be responsible for large and stable confidence biases, namely limited discrimination, the hard-easy effect, the Dunning-Kruger effect, conservative learning from experience and the overprecision phenomenon (without underprecision) if subjects act as Bayesian learners who rely only on sequentially perceived performance cues and contrarian illusory signals induced by doubt. Moreover, these biases are likely to persist since the Bayesian aggregation of past information consolidates the accumulation of errors and the perception of contrarian illusory signals generates conservatism and under-reaction to events. Taken together, these two features may explain why intuitive Bayesians make systematically wrong predictions of their own performance.

Chapter 2 aims at understanding the impact of creativity on economic outcomes. The first goal of this chapter is to review how economists describe creative behavior and propose how it should be described. We argue that from an economic perspective, creative behavior must be judged by individual's propensity to innovate in production (and consumption) activities, distinguishing two types of economic innovators: *researchers* (the ability to find new solutions) and *entrepreneurs* (the ability to capture unexpected rents). The second goal is to observe how the potential of creativity impacts individuals' production. We propose an economic experiment with two real-effort tasks to observe the performance of creative individuals in production, using three psychological measures of creativity: the graphical divergent thinking, the graphical integrative thinking and an aggregated creativity index. We find that divergent thinking correlates with the *researcher* type of economic innovator since higher scores for this psychological measure of creativity increase the productivity in exploration activities. Otherwise, the *entrepreneurial* type was not identified among our creativity scores. Additionally, we observe that creative individuals are no more productive than others in repetitive tasks, but they behave differently than less creative individuals in this type of task:

integrative thinkers are more cooperative when working in pairs, maybe because they are intelligent and understand the benefit of cooperation in teamwork. Indeed, the idea that creative individuals are intelligent is reinforced by the performance at school -a real life performance. Creativity scores play an important role on school achievements, they are positively correlated to grades on Maths, French and the general grade.

Chapters 3 and 4 are based on the same experiment and dataset, with a total of 941 participants. We reproduce experimentally the typical structure of schooling systems, and propose a real-effort task (solving anagrams). After a long phase of compulsory schooling (level 1), students may quit for the job market or engage in further studies. Those who decide to continue usually have an option between two tracks (or more), a general and a vocational track, which differ in the required level of cognitive ability. The less able students should opt for vocational studies in level 2 while the more able would opt for general studies. If successful, both groups of students would have another choice to quit or engage in further studies (level 3). However, students engaged in general education would normally find it a lot easier to pass this higher level than students engaged in a vocational track. We compare four mechanisms for sorting students according to their abilities: self-selection of further studies with a single track (no-choice of track), self-selection of further studies with the choice of track (choice), screening by ability and early *numerus clausus* competition.

Chapter 3 shows that No-choice and Screening are the more efficient mechanisms, providing higher payoffs, outcomes and a higher rate of success at tertiary level. Screening results in the highest output (number of solved anagrams) for the primary level as it stimulates sustained effort of individuals at this level. Early competition (Race) is the worst treatment because participants care not only about their own performance but also about others' performance. The problem of self-selection (Choice) is that it promotes the highest level of failure at secondary level when the economic returns to school are high. In fact, we observe that higher returns to tertiary education increase

the amount of regret. Thus, we observe that the inefficiency of the system derives from two main reasons: (i) if students have an imperfect knowledge of their own ability, and/or if there is a lack of discrimination between the two educational tracks (Vocational and General), they are inclined to opt for the more difficult track and fail, reducing the wealth generated in the experiment by 12%; (ii) the higher is the wage premium for tertiary education the higher is the ex-ante expected utility, what raises the chances to try higher levels of education, but does not increase the probability of success, increasing the ex-post inefficiency (higher level of failures, dropouts and regret).

The question studied in chapter 4 is: *how do different school systems and school returns differently affect ability groups, genders, and social groups, thus causing substantial differences in social and gender bias among developed countries and periods?* We find that competition is the worst institution for high and medium ability individuals, while self-selection of track is the worst treatment for low-ability individuals when returns to tertiary education are high. The main result observed when increasing payoffs for the tertiary level is that a rise in tertiary education is beneficial for high and medium-ability students, but it is harmful to low-ability students. Comparing payoffs for the two choice treatments seems unfair at first, but surprisingly, low ability participants earned 22% more in the condition with lower incentives. This effect is due to better decisions.

Thus, since competition is especially harmful to high and medium ability individuals, it appears to cause lower socioeconomic bias than other sorting mechanisms, a direct consequence of the relative inefficiency of this mechanism. The impact of Choice with high incentives over the performance of less able participants has the contrary effect, it generates the highest socioeconomic bias among our treatments. We observe also that the random allocation is the only mechanism that is fair for gender differences. Screening seems to be the most balanced mechanism to track students by ability, the

challenge is to set fair grades (thresholds) that encourage motivated low and medium ability students to reach higher levels of education, without discouraging the less motivated to complete the primary level. An equitable system should improve the outcomes of less able individuals, without prejudice for the more able.

Part I

Self-confidence and Creativity

CHAPTER 1

CONFIDENCE BIASES AND LEARNING AMONG
INTUITIVE BAYESIANS

This chapter is a joint work with Louis Lévy-Garboua and Muniza Askari. Published in *Theory and Decision*.

1. Introduction

In many circumstances, people appear to be "overconfident" in their own abilities and good fortune. This may occur when they compare themselves with others, massively finding themselves "better-than-average" in familiar domains (eg., [Svenson 1981](#), [Kruger 1999](#)), when they overestimate their own absolute ability to perform a task (eg., [Lichtenstein and Fischhoff 1977](#), [Lichtenstein, Fischhoff and Phillips 1982](#)), or when they overestimate the precision of their estimates and forecasts (eg., [Oskamp 1965](#)). [Moore and Healy \(2008\)](#) designate these three forms of overconfidence respectively as overplacement, overestimation, and overprecision. We shall here be concerned with how people overestimate, or sometimes underestimate, their own absolute ability to perform a task in isolation. Remarkably, however, our explanation of the estimation bias predicts the overprecision phenomenon as well.

The estimation bias refers to the discrepancy between ex post objective performance (measured by frequency of success in a task) with ex ante subjectively held confidence ([Lichtenstein, Fischhoff and Phillips, 1982](#)). It has first been interpreted as a cognitive bias caused by the difficulty of the task (e.g., [Griffin and Tversky 1992](#)). It is the so called "hard-easy effect" ([Lichtenstein and Fischhoff, 1977](#)): people underestimate their ability to perform an easy task and overestimate their ability to perform a difficult task. However, a recent literature has challenged this interpretation by seeking to explain the apparent over/underconfidence by the rational-Bayesian calculus of individuals discovering their own ability through experience and learning ([Moore and Healy, 2008](#), [Grieco and Hogarth, 2009](#), [Benoît and Dubra, 2011](#), [Van den Steen, 2011](#)). While the cognitive bias view describes self-confidence as a stable trait, the Bayesian learning perspective points at the experiences leading to over- or under-confidence. The primary goal of this paper is to propose a parsimonious integration of the cognitive bias and the learning approach.

We design a real-effort experiment which enables us to test the respective strengths of estimation biases and learning. People enter a game in which the task becomes increasingly difficult -*i.e.* risky- over time. By comparing, for three levels of difficulty, the subjective probability of success (confidence) with the objective frequency at three moments before and during the task, we examine the speed of learning one's ability for this task and the persistence of overconfidence with experience. We conjecture that subjects will be first underconfident when the task is easy and become overconfident when the task is getting difficult. However, "difficulty" is a relative notion and a task that a low-ability individual finds difficult may look easy to a high-ability person. Thus, we should observe that overconfidence declines with ability and rises with difficulty. The question raised here is the following: if people have initially an imperfect knowledge of their ability and miscalibrate their estimates, will their rising overconfidence as the task becomes increasingly difficult be offset by learning, and will they learn their true ability fast enough to stop the game before it is too late?

The popular game "*double or quits*" fits the previous description and will thus inspire the following experiment. A modern version of this game is the world-famous TV show "*who wants to be a millionaire*". In the games of "*double or quits*" and "*who wants to be a millionaire*", players are first given a number of easy questions to answer so that most of them win a small prize. At this point, they have an option to quit with their prize or double by pursuing the game and answering a few more questions of increasing difficulty. The same sort of double or quits decision may be repeated several times in order to allow enormous gains in case of repeated success. However, if the player fails to answer one question, she must step out of the game with a consolation prize of lower value than the prize that she had previously declined.

Our experimental data reproduces the double or quits game. We observe that subjects are under-confident in front of a novel but easy task, whereas they feel over-confident and willing to engage in tasks of increasing difficulty to the point of failing.

We propose a new model of "intuitive Bayesian learning" to interpret the data and draw new testable implications. Our model builds on ideas put forward by Erev, Wallsten and Budescu (1994) and Moore and Healy (2008). It is Bayesian like Moore and Healy (2008), while viewing confidence as a subjective probability of success, like Erev, Wallsten and Budescu (1994). However, it introduces intuitive rationality to overcome a limitation of the rational-Bayesian framework which is to describe how rational people learn from experience without being able to predict the formation of confidence biases *before* completion of a task. This is not an innocuous limitation because it means, among other things, that the rational-Bayesian theory is inconsistent with the systematic probability distortions observed in decisions under risk or uncertainty since the advent of prospect theory (Kahneman and Tversky, 1979). Therefore, we need to go deeper into the cognitive process of decision. Subjects in our view derive their beliefs exclusively from their prior and the informative signals that they receive. However, "intuitive Bayesians" decide on the basis of the sensory evidence that they perceive sequentially. If they feel uncertain of their prior belief, they will perceive the objection to it triggered by their doubt and wish to "test" its strength before making their decision, like those decision makers weighting the pros and cons of an option. The perceived objection to a rational prior acts like a *contrarian illusory signal* that causes probability distortions in opposition to the prior and this is a cognitive mechanism that does not require completion of the task. As they gain experience, they keep on applying Bayes rule to update their prior belief both by cues on their current performance and by the prior-dependent contrarian signal. Thus, with the single assumption of intuitive rationality, we can account for all the cognitive biases described on our data within the Bayesian paradigm and integrate the cognitive bias and the learning approach. With this model, and in contrast with Gervais and Odean (2001), we don't need to assume a self-attribution bias (Langer and Roth, 1975, Miller and Ross, 1975)

combined with Bayesian learning to produce overconfidence¹. Signals of future success and failure are treated symmetrically². Finally, unlike models of confidence management (e.g. Brunnermeier and Parker 2005, Köszegi 2006, Möbius et al. 2014), we don't have to postulate that individuals manipulate their beliefs and derive direct utility from optimistic beliefs about themselves.

Section 2 lays down the structure of the experiment and incentives, and provides the basic descriptive statistics. Our large data set allows a thorough description of confidence biases and a dynamic view of their evolution with experience of the task. Section 3 describes the confidence biases and learning shown by our data. Four basic facts about confidence are reported from our data: (i) limited discrimination among different tasks; (ii) miscalibration of subjective probabilities of success elicited by the "hard-easy effect"; (iii) differential, ability-dependent, calibration biases known as the Dunning-Kruger (or ability) effect (Kruger and Dunning, 1999); and (iv) local, but not global, learning. Section 4 proposes a new theory of over (under)-confidence among intuitive Bayesians which integrates doubt and learning and can predict biases, before as well as during the task, in repeated as well as in single trials. Doubt-driven miscalibration appears to be a sufficient explanation, not only for the hard-easy effect and the 'ability' or Dunning-Kruger effect, but also for limited discrimination and for the over-precision phenomenon. The theory is further used in section 5 to predict the evolution of confidence over experience on our data set. For instance, low-ability subjects first lose confidence when they discover their low performance during the first and easiest level; but they eventually regain their initial confidence in own ability to perform more difficult tasks in the future after laborious but successful completion of the first level.

1. Using German survey data about stock market forecasters, Deaves, Lüders and Schröder (2010) does not confirm that success has a greater impact than failure on self-confidence, which casts doubt on the self-attribution bias explanation.

2. In studies where subjects are free to stay or to leave after a negative feedback, subjects who update most their confidence in their future success to a negative feedback are selectively sorted out of the sample. This creates an asymmetry in measured responses to positive and negative feedback. Such spurious asymmetry does not exist in the present experiment, because subjects who fail to reach one level must drop out of the game.

Intuitive Bayesians exhibit *conservatism*, that is, under-reaction to received information, and slow learning. Finally, we show in sub-section 5.3 that the cues upon which subjects construct their own estimate of success, *i.e.* confidence, widely differ from the genuine predictors of success, which further explains the *planning fallacy*³. The conclusion follows in section 6.

2. The experiment

2.1. Task and treatments

Participants perform a real-effort, rather long and difficult, task for which they get paid according to their degree of success. The task consists in solving anagrams ranked in three levels of increasing difficulty. It is performed during a maximum of 15 rounds lasting no more than 8 minutes each. These 15 rounds are structured in three successive levels of increasing difficulty, designated respectively as the training level, the middle level, and the high level.

Participants are successful at one level when they manage to decode 2/3 of the anagrams at this level. An example of the task screen is reproduced in appendix. The training level consists of 9 rounds of low difficulty (*i.e.* 6 anagrams per round to be solved in no more than eight minutes). It is long enough to let participants feel that a large effort and ability is required of them to succeed at the optional upper levels. It does also let them ample time to learn the task. The middle and high levels, which come next, comprise 3 rounds each.

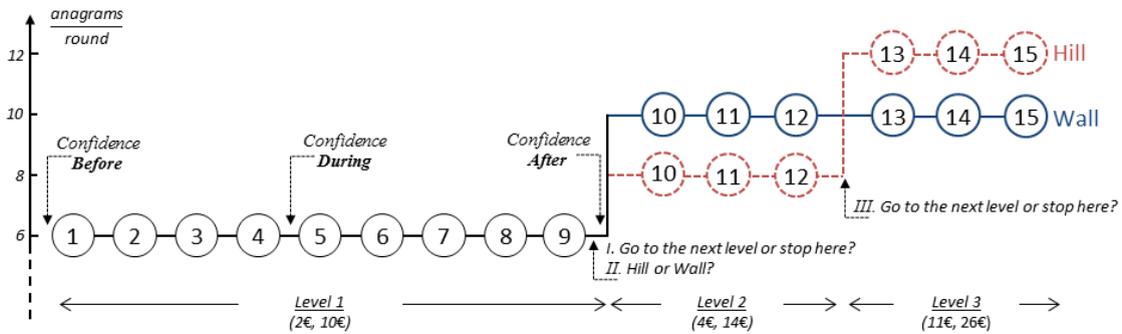
The gradient of task difficulty was manipulated after completion of the training level and two conditions are available: (*i*) in the '*wall*' condition, the difficulty jumps sharply at middle level, but remains constant at high level; (*ii*) in the '*hill*' condition, the difficulty always rises from one level to the next, slowly first at middle level, then

3. The planning fallacy is the tendency to underestimate the time needed for completion of a task. See, e.g. Buehler, Griffin and Ross (2002).

sharply at high level.

By the end of the experiment, the required number of anagrams is the same for the 'wall' and 'hill' conditions. However, the distribution of anagrams to be decoded differs for these two conditions. In the wall condition, ten anagrams per round are proposed at the middle and high levels, of which 20 anagrams at least must be decoded per level. In the hill condition, eight anagrams per round are proposed at middle level, and this rises to twelve anagrams at high level. Decoding sixteen anagrams in three rounds is required for middle level; and decoding twenty-four anagrams in three rounds is required for high level. This design can be visualized in Figure 1.1. The same figure appears (without the legends) on the screen before each round ⁴.

Figure 1.1 – Decision problem perceived by participants at the start of level 2 of the choice treatment.



Notes:

Payoffs in parentheses : (fail, success and stop).

Decisions I, II and III are conditioned to success in the previous level.

Decision II depends on the treatment.

Estimation of Confidence After is conditioned to success in the first and decision to start the second level.

The manipulation of the 'wall' and 'hill' conditions gave rise to three treatments:

- Wall treatment (*wall*): the wall condition is imposed to participants who passed the training level;

4. The screen highlights the round, the number of correct anagrams cumulated during the current level and the number of anagrams needed to pass this level.

- Hill treatment (*hill*): the hill condition is imposed to participants who passed the training level;
- Choice treatment (*choice*): a choice among the two conditions (*wall* or *hill*) is proposed to participants who passed the training level.

The double or quits game is played under these three treatments. All subjects first go through the training level. Those who were successful -i.e., those who solved at least 36 anagrams during the training level- will then be asked to double or quits:

- **Double**: Continue to the next level to win a substantial increase in earnings;
- **Quits**: Stop the experiment and take your earnings.

Participants who decide to go to middle level get a consolation prize that is lower than the foregone earnings if they fail or drop out before the third round. If they succeed middle level, they will be asked again to double or quits. The same rules apply for high level at rising levels of earnings. The potential gains (in Euros) were (10, 2) at the training level, that is, 10€ for successful quitters and 2€ for failures, (14, 4) at middle level, and (26, 11) at high level.

2.2. Experimental sessions

We ran 24 sessions for a total of 410 participants, half for the choice treatment and the other half equally split between the '*wall*' and '*hill*' treatments. Eight sessions were run in the BULCIRANO lab (Center for Interuniversity Research and Analysis on Organizations), Montreal (Canada), and the same number of sessions were conducted at the LEEP (*Laboratoire d'Economie Expérimentale de Paris*), Pantheon-Sorbonne University. The difference between Paris and Montreal was observed to be insignificant. Thus, eight additional sessions were conducted at LEEP in order to acquire robust results. A show-up fee of 5€ in Paris and Can\$ 5 in Montreal was paid to the participants (from now on, all money amounts will be given in Euros). About 80% of the participants were students.

At the start, instructions were read out and a hard copy of it was also provided individually. Participants answered six questions to test their full comprehension of the experiment. Information on gender, age, educational level and labor market status was required. The last question was a hypothetical choice between 5€ for sure and an ambiguous urn containing 100 balls of two colors (white and black) in unknown proportions. Ten Euros (10€) were to be earned if a black ball was drawn. Choice of the sure gain provided a rough but simple measure of risk aversion in the uncertainty context of the experiment.

2.3. Descriptive statistics

The main descriptive statistics for the three treatments are reported in Table 1.1:

Table 1.1 – Descriptive statistics for the three treatments

Variables	Treatments		
	Wall	Hill	Choice
Male	56%	48%	49%
Age	24.5	25.8	25.1
Risk Averse	54%	59%	51%
Payments	9.1	8.9	7.8
Total anagrams solved	55.6	53.7	54.3
Ability ⁵	2.8	2.7	2.6
Number of observations	101	106	203
<i>Decision to double conditional on success at previous level:</i>			
Middle level	78% (91)	76% (90)	77% (176)
High level	95% (22)	72% (29)	82% (34)

Notes: Decision to double to High level: difference between the "Wall" and "Hill" treatments is significant at 5%; all other differences are not significant at 10% level (t-test). Number of participants successfully clearing the previous level is in parentheses.

The results of tests show that the three samples are homogeneous. No significant difference is observed among the samples' means for individual characteristics. As expected, the 'wall' and 'hill' treatments had a substantial impact on the decision to double upon reaching the middle level. Almost everybody doubles in the 'wall'

5. Ability is measured by the number of anagrams solved per minute in the first 4 rounds. It lies in the interval [0,6].

treatment on reaching middle level because the high level is no more difficult than the middle level. In contrast, only 72% enter the high level in the 'hill' treatment as the difficulty gradient is very steep (t-test: $t= 2.20$; $p\text{-value}=0.033$). In spite of these differences, the number of anagrams solved and payments may be considered equal among treatments at the usual level of significance.

Subjects can also be grouped in three different levels of ability, according to the number of anagrams solved per minute in the first 4 rounds: high ability (first tercile), medium ability (second tercile) and low ability (last tercile). Some descriptive statistics for the three treatments are reported on Table 1.2. The three groups are homogeneous in terms of gender and risk aversion but a slightly greater proportion of low-ability subjects can be found among older, probably non-student, participants.

Table 1.2 – Descriptive statistics by ability level

Variables	Level of ability			Difference		
	High	Medium	Low	M-H	L-M	L-H
Male	47%	54%	50%	<i>ns</i>	<i>ns</i>	<i>ns</i>
Age	23.6	24.5	27.2	<i>ns</i>	***	***
Risk Aversion	53%	50%	59%	<i>ns</i>	<i>ns</i>	<i>ns</i>
Payments	11.7	7.7	6.0	***	**	***
Number anagrams solved	67.7	53.8	42.6	***	***	***
Ability	4.5	2.4	1.1	***	***	***
Number of observations	131	142	137			
<i>Decision to double conditional on success at previous level:</i>						
Middle level	91% (128)	81% (127)	54% (102)	**	***	***
High level	87% (55)	72% (25)	80% (5)	*	<i>ns</i>	<i>ns</i>

Notes: Significance level: * 10%; ** 5%; ***1%; ns: not significant at 10% level (t-test). Number of participants successfully clearing the previous level is in parentheses

Table 1.2 shows that "ability" strongly discriminates among participants in terms of performance (total anagrams solved, payments) and quits before the middle level. However, the training level was meant to be easy enough that three-quarters (102:137) of low-ability subjects would pass it.

2.4. Confidence judgments

Participants were asked to state their subjective probability of success for the three levels and at three moments: before, during, and after the training level. Before beginning the game, they were shown a demonstration slide which lasted one minute. Anagrams of the kind they would have to solve appeared on the screen with their solution. Then, they were asked to assess their chances of success on a scale of 0 to 100 (Adams, 1957), and the game started for real. After four rounds of decoding anagrams, players were asked again to rate their confidence. Lastly, players who had passed the training level and decided to double re-estimated their chances of success for the middle and high levels.

The Adams's (1957) scale that we used is convenient for quantitative analysis because it converts confidence into (almost) continuous subjective probabilities. It was required for consistency that the reported chances of success do not increase as the difficulty level increased. Answers could not be validated as long as they remained inconsistent. Subjects actually used the whole scale but, before the experiment, 14% expressed absolute certainty that they would succeed the first level and only 1 participant was sure that she would fail.

We did not directly incentivize beliefs because our primary aim was not to force subjects to make optimal forecasts of their chances of success but to have them report sincerely their true beliefs in their attempt to maximize their subjective expected utility, and to observe the variation of such beliefs with experience. The true beliefs are those which dictate actual behavior following such prediction, and the latter was incentivized by the money gains based on subjects' decisions to double or quits and performance in the task. Armantier and Treich (2013) have recently generalized previous work on proper scoring rules (see their extensive bibliography). They show that, when subjects have a financial stake in the events they are predicting and can hedge their predictions by taking additional action after reporting their beliefs, use of any

proper scoring rule generates complex distortions in the predictions and further behavior since these are not independent and are in general different from what they would have been if each had been decided separately. In the present context, final performance yields income and does not immediately follow the forecast. Hence, incentivizing forecasts might force subjects to try and adjust gradually their behavior to their forecast and, therefore, unduly condition their behavior. A further difficulty encountered in this experiment was that, by incentivizing beliefs on three successive occasions, we induced risk-averse subjects to diversify their reported estimates as a hedge against the risk of prediction error. Self-report methods have been widely used and validated by psychologists and neuroscientists; and recent careful comparisons of this method with the quadratic scoring rule⁶ found that it performed as well (Clark and Friesen, 2009) or better (Hollard, Massoni and Vergnaud, 2015) than the quadratic scoring rule⁷. Considering that self-reports perform nicely while being much simpler and faster than incentive-compatible rules, use of the self-report seemed appropriate in this experiment.

3. Describing confidence biases and learning

3.1. Limited Discrimination

About half of our subjects were selected randomly into the 'wall' and 'hill' treatments and could not choose between the two. Those selected in one path were informed of the characteristics of their own path but had no knowledge whatsoever of the char-

6. After the subject has reported a probability p , the quadratic scoring rule imposes a cost that is proportional to $(1 - p)^2$ in case of success and to $(0 - p)^2$ in case of failure. The score takes the general form: $S = a - b$. Cost, with $a, b > 0$.

7. The second study also included the lottery rule in the comparison and found that the latter slightly outperformed self-report. The lottery rule rests on the following mechanism: after the subject has reported a probability p , a random number q is drawn. If q is smaller than p , the subject is paid according to the task. If q is greater than p , the subject is paid according to a risky bet that provides the same reward with probability q . The lottery rule cannot be implemented on our design.

acteristics, nor even the existence, of the other path.

Result 1 (Limited discrimination): *Subjects do not perceive differences of difficulty between two different tasks in the future unless such differences are particularly salient. Moreover, they are not forward-looking, in the sense that they are unable to anticipate the increased likelihood of their success at the high level conditional on passing the middle level. However, they can be sophisticated when it is time for them to choose.*

Support of result 1: Table 1.3 compares confidence judgments regarding the three levels of difficulty among the 'wall' and the 'hill' subjects before, during, and after the training period. Although the 'wall' and 'hill' were designed to be quite different at the middle and high levels, the subjective estimates of success exhibit almost no significant difference at any level. The single exception concerns the early estimate (before round 1) regarding the high level for which the difference of gradient between the two paths is particularly salient. However, the difference ceases to be significant as subjects acquire experience of the task. This striking observation suggests that individuals are unable to discriminate distinctive characteristics of the task unless the latter are particularly salient.

Perhaps even more disturbing is the fact that, in Table 1.3, subjects discount their confidence level from the middle to the high level as much in the Wall as in the Hill treatment. For instance, just before the middle level, the ratio of confidence in passing the high level to confidence in passing the middle level was close to 0.70 in both treatments. However, a perfectly rational agent should realize that the high level is no more difficult than the middle level in the Wall treatment whereas it is much more difficult in the Hill treatment. Thus, she should report almost the same confidence at both levels in the Wall treatment, and a considerably lower confidence at the high

level in the Hill treatment. The latter observation suggests that most individuals are unable to compute conditional probabilities accurately even when the latter is equal to one as in the Wall treatment. They don't anticipate that, if they demonstrate the ability to solve 20 anagrams or more at middle level, they should be almost sure to solve 20 or more at the high level. However, subjects do make the right inference when it is time for them to make the decision since 95% of subjects who passed the middle level in the Wall treatment decided to continue (Table 1.1). And, if they have a choice between Wall and Hill, they do make a difference between these two tracks: 71.4% of doublers then prefer the Wall track although they would have greater chances of success at the middle level if they chose Hill. This observation suggests that subjects did not maximize their immediate probability of success but made a sophisticated comparison of the expected utility of both tracks, taking the option value of Wall in consideration before making an irreversible choice of track spanning over two periods⁸.

Table 1.3 – A comparison of confidence for the wall and hill treatments shown separately

Subjective confidence	No-choice treatment		Difference	
	Wall (%)	Hill (%)		
Before round 1:	<i>Level 1</i>	80	77	<i>ns</i>
	<i>Level 2</i>	62	58	<i>ns</i>
	<i>Level 3</i>	47	40	**
Before round 5:	<i>Level 1</i>	71	71	<i>ns</i>
	<i>Level 2</i>	53	52	<i>ns</i>
	<i>Level 3</i>	40	36	<i>ns</i>
Before round 10:	<i>Level 2</i>	60	56	<i>ns</i>
	<i>Level 3</i>	43	39	<i>ns</i>

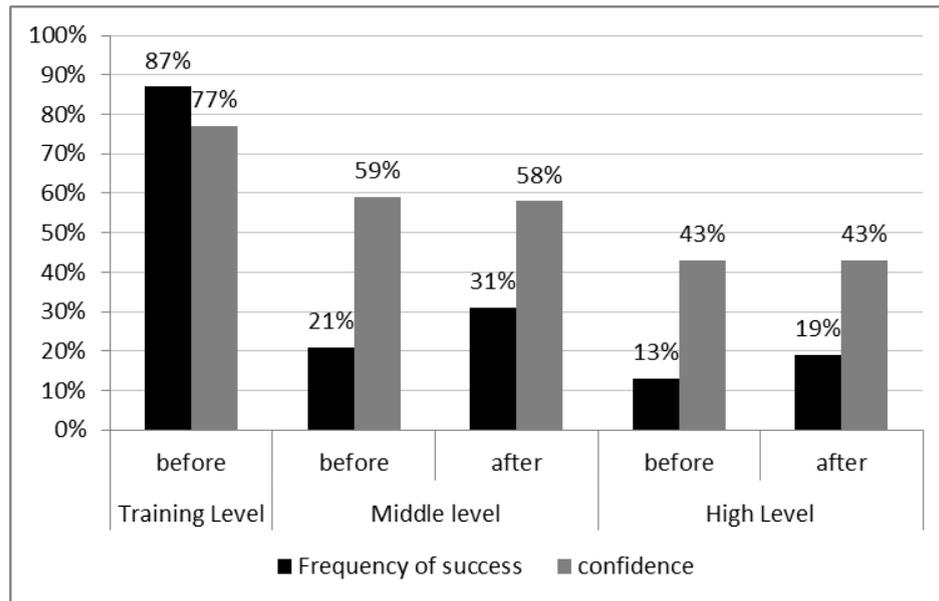
Notes. Observations: Before rounds 1 and 5 (before round 10): 101 (71) for wall and 106 (68) for hill. **Significance Level:** * $p < .10$, ** $p < .05$, *** $p < .01$, ns: not significant at 10% level.

3.2. Miscalibration

Result 2 (The hard-easy effect): *In comparison with actual performance, confidence in one's ability to reach a given level is underestimated for a novel but relatively*

8. We are grateful to Luis Santos-Pinto for making the last point clear in early discussions.

Figure 1.2 – Hard-easy effect observed at three levels



Notes: **Observations:** before training level (N: 410); after training level (N: 275 - analysis restricted to doublers). **Differences** between frequency of success and confidence (before and after) are significant at 1% at all levels (Training, Middle and High). (*t-test*)

easy task (the training level); and it is overestimated for the subsequent more difficult tasks (the middle and high levels). Overconfidence increases in relative terms with the difficulty of the task. Conditional on an initial success (training level) and on the decision to continue, confidence in one's ability to reach higher levels is still overestimated. Thus, initially successful subjects remain too optimistic about their future.

Support for result 2: Figure 1.2 compares the measured frequency of success with the reported subjective confidence in the three successive levels of increasing difficulty. For the middle and high levels, we also indicate these two probabilities as they appear before the training period and after it conditional on doubling. The Choice and No-choice conditions have been aggregated on this figure because no significant difference was found in the result of tests.

The task required at the training level was relatively easy for our subjects since 87% passed this level. However, subjects started it without knowing what it would be like and, even after four rounds of training, they underestimated their own ability to a low 77% probability of success. The difference among the two percentages is significant ($t=5.77$, $p=0.000$; t-test). Hence, individuals are under-confident on the novel but relatively easy task.

In contrast, subjects appear to be overconfident as the task gets increasingly difficult. They consistently diminish their estimated probabilities of success but do not adjust their estimates in proportion to the difficulty of the task. Thus, individuals tend to overestimate their own chances for the advanced levels. The difference between the frequency of success and confidence before the task is always significant, both at the middle level ($t=18.3$, $p=0.000$) and at the high level ($t=17.1$, $p=0.000$).

The same conclusions hold conditional on passing the training level and choosing to double. Subjects remain overconfident in their future chances of success. However, their confidence does not rise after their initial success in proportion to their chances of further success.

3.3. The ability effect

Result 3 (The ability effect): *Overcalibration diminishes with task-specific ability.*

Support for result 3: The hard-easy effect is reproduced on Figures 1.3a, 1.3b, 1.3c for the three ability terciles⁹. Low-ability subjects are obviously more overconfident at middle and high levels relative to high and medium-ability individuals. This result confirms earlier observations of Kruger and Dunning (1999) among others (see Ryvkin, Krajč and Ortmann (2012) for a recent overview and incentivized experiments). The so-called Dunning-Kruger effect has been attributed to a metacognitive inability of the

9. Difference between confidence and frequency of success is significant at 1% for all ability levels. For these figures, we selected confidence reported after 4th round (during training level) in order to minimize the impact of mismeasurement.

Figure 1.3a – Under-confidence at the training level, by ability.

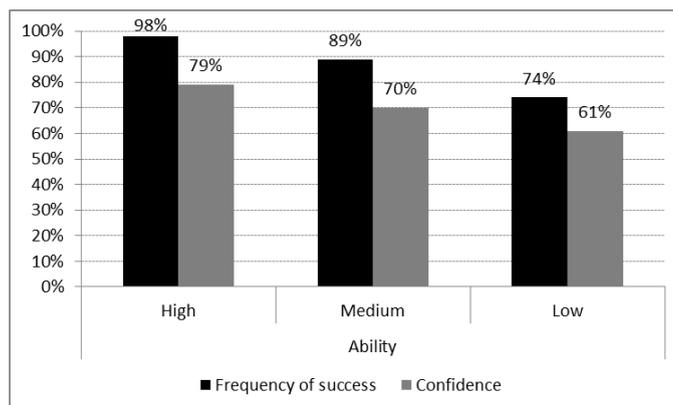
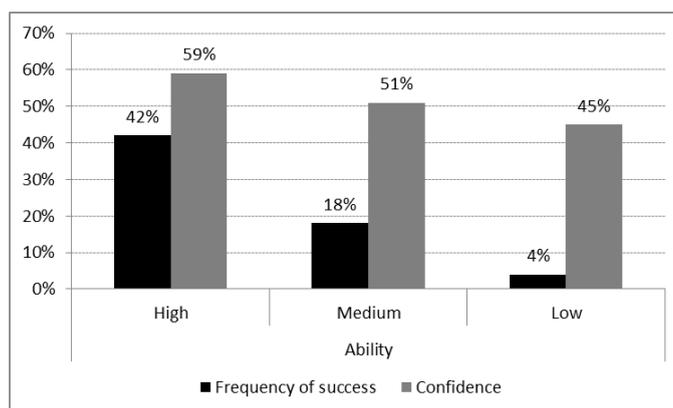


Figure 1.3b – Overconfidence at middle level, by ability.

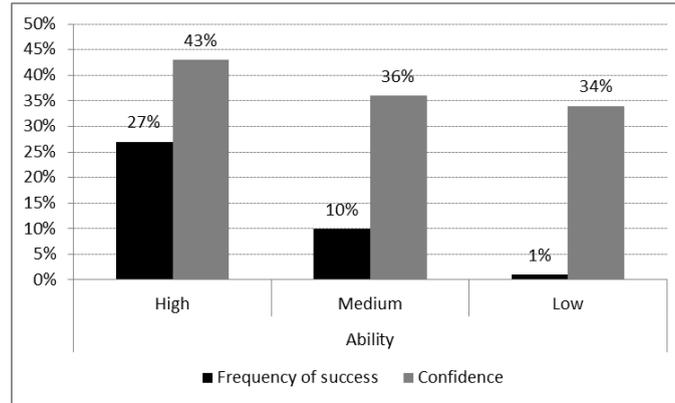


unskilled to recognize their mistakes¹⁰. We give here another, and in our opinion, simpler explanation¹¹. The ability (or Dunning-Kruger) effect may be seen as a corollary of the hard-easy effect because "difficulty" is a relative notion and a task that a low-ability individual finds difficult certainly looks easier to a high-ability person. Thus, if overconfidence rises with the difficulty of a task, it is natural to observe that it declines on a given task with the ability of performers.

10. The Dunning-Kruger effect initially addressed general knowledge questions whereas we consider self-assessments of own performance in a real-effort task.

11. Our explanation may also be better than the initial explanation such that the unskilled are unaware of their lower abilities. Miller and Geraci (2011) found that students with poor abilities showed greater overconfidence than high-performing students, but they also reported lower confidence in these predictions.

Figure 1.3c – Overconfidence at high level, by ability.



3.4. Learning

Result 4 (Learning is local, not global): *Confidence and performance co-vary during the task. Subjects learned **locally** upon experiencing variations in their performance. However, they didn't learn **globally** in our experiment, since doublers remained as confident as before after completing the training level irrespective of their true ability level.*

Support for result 4: Figures 1.4 and 1.5 describe confidence by ability group before, during, and after the training period^{12 13} for the middle and high level respectively whereas Figure 1.6 describes the variation of performance of the same groups within the same period. These graphs, taken together, show a decline in both (ability-adjusted) confidence and performance during the first four rounds, followed by a concomitant

12. No significant difference was found between the Choice and No-choice conditions, suggesting that the option to choose the preferred path does not trigger an illusion of control.

13. Participants who reported confidence after the training period were more able than average since they had passed this level and decided to double. Thus, we compare ability-adjusted confidence Before and During with the reported confidence After. The ability-adjusted confidence Before and During are obtained by running a simple linear regression of confidence Before and During on ability, measured by the average number of anagrams solved per minute in the first 4 rounds of the training level. The estimated effect of superior ability of doublers was added to confidence During or Before to get the ability-adjusted confidence which directly compares with the observed confidence After.

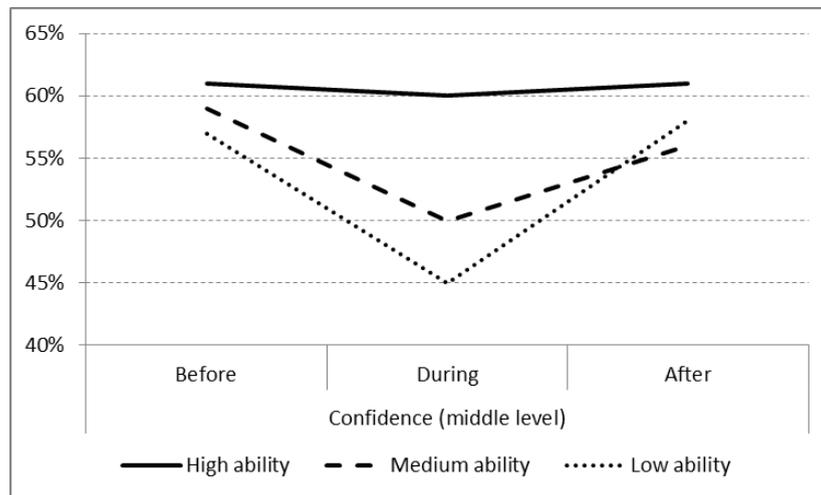
rise of confidence and performance in the following rounds¹⁴. The observed decline of confidence at the beginning of the training period can be related on Figure 1.6 to the fact that participants solved less and less anagrams per period during the first four periods: 5.51 on average in period 1, 5.18 in period 2, 4.60 in period 3, and 4.17 in period 4¹⁵. Subjects kept solving at least two-thirds of the anagrams available during the training session but probably lost part of their motivation on repeating the task. On sequentially observing their declining performance, they revised their initial estimate of future success downward. However, on being asked to report their confidence after four rounds, they became conscious of their performance decline and responded to this information feedback. Performance rose sharply but momentarily during the next two rounds. The average performance first rose to 4.37 in period 5 and 5.05 in period 6 then sharply declined to 4.39 in period 7, 4.06 in period 8 and 3.48 in period 9. As soon as subjects became (almost) sure of passing the training level, they diminished their effort. During the experiment it was also observed that individuals stopped decoding further anagrams as soon as the minimum requirement to clear a level was fulfilled.

Subjects experiencing low (medium) performance in the first rounds seem to learn locally that they have a low (medium) ability since the confidence gap widens during the first four periods. However, this learning effect is short-lived since the confidence gap shrinks back to its initial size after low (medium)-ability subjects strove to succeed, increasing their performance (as reported on Figure 1.6) and regaining confidence. Eventually, experienced "doubblers" are as confident to succeed at higher levels as they were before the task, irrespective of their ability level: there is no global learning effect. We share the conclusion of Merkle and Weber (2011) that the persistence of prior beliefs is inconsistent with fully rational-Bayesian behavior (see also Benoît, Dubra and Moore 2015).

14. With a single exception, confidence variations are statistically significant at 1% level in the middle and high levels.

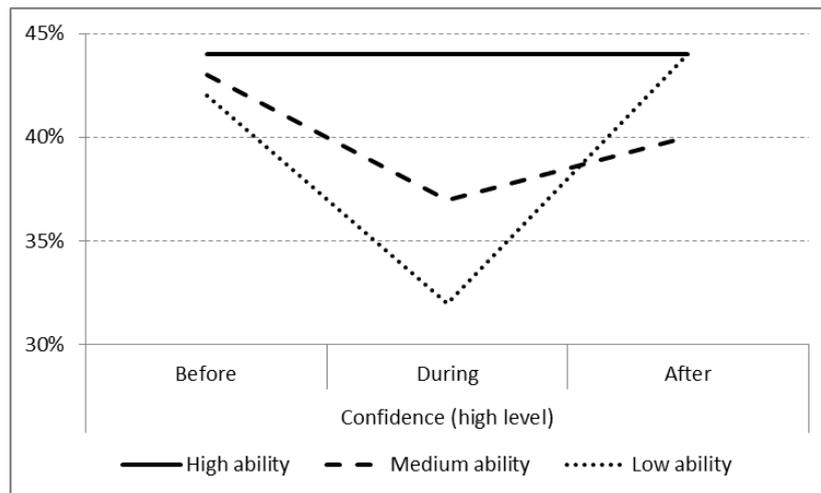
15. There was no significant difference between treatments.

Figure 1.4 – Variation of confidence with experience, by level of ability: middle level



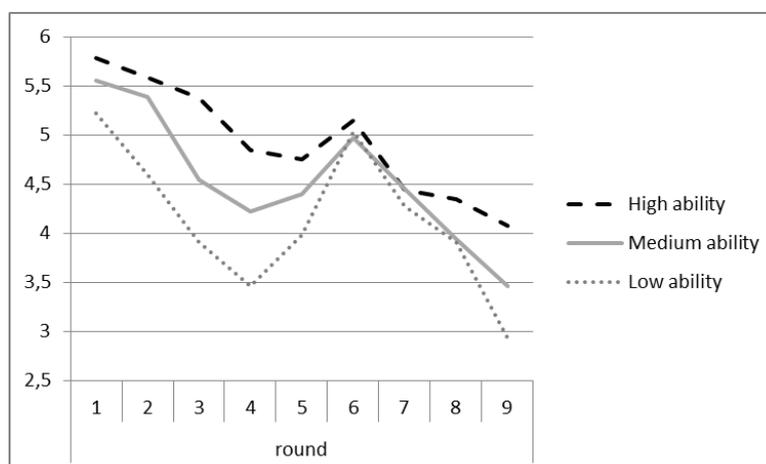
Notes. **Sample size:** 410 individuals for Before and During, and 275 for After (only doublers). We report the adjusted ability for doublers, see Footnote 13 for more details. **Differences between ability levels** are significant at 1% level Before and During. Differences After are not significant at 10% level. **Differences by ability level:** High-ability: During-Before: ***; After-During: ns; After-Before: ns. Medium-ability: During-Before: ***; After-During: ***; After-Before: **. Low-ability: During-Before: ***; After-During: *** ; After-Before: ns. **Significance level:** *** 1%; ** 5%; * 10%; ns: not significant at 10% level (t-test).

Figure 1.5 – Variation of confidence with experience, by level of ability: high level



Notes. **Sample size:** 410 individuals for Before and During, and 275 for After (only doublers). We report the adjusted ability for doublers, see Footnote 13 for more details. **Differences between ability levels** are significant at 1% level Before and During. Differences After are not significant at 10% level. **Differences by ability level:** High-ability: During-Before: ns; After-During: ns; After-Before: ns. Medium-ability: During-Before: ***; After-During: **; After-Before: ns. Low-ability: During-Before:***; After-During: *** ; After-Before: ns.. **Significance level:** *** 1%; ** 5%; * 10%; ns: not significant at 10% level (t-test).

Figure 1.6 – Number of anagrams solved per round by level of ability



4. Theory

We present now a simple Bayesian model that describes absolute confidence reported before and during completion of a task, and predicts limited discrimination, the hard-easy effect and the ability effect. It builds on ideas put forward by [Erev, Wallsten and Budescu \(1994\)](#) and [Moore and Healy \(2008\)](#) who both consider that confidence, like most judgments, are subject to errors. [Erev, Wallsten and Budescu \(1994\)](#) view confidence as a subjective probability that must lie between 0 and 1. Hence, probabilities close to 1 are most likely to be underestimated and probabilities close to 0 are most likely to be overestimated. The hard-easy effect and the ability effect may be merely the consequence of that simple truth. However, their theory offers a qualitative assessment that lacks precision and cannot be applied to intermediate values of confidence. [Moore and Healy \(2008\)](#) analyze confidence as a score in a quiz that the player must guess *after* completion of the task and *before* knowing her true performance. Bayesian players adjust their prior estimate after receiving a subjective signal from their own experience. It is natural to think that signals are randomly distributed around their true unknown value. Assuming normal distributions for the signal and the prior, the posterior expectation of confidence is then a weighted average of the prior and the signal lying necessarily between these two values. Thus, if the task was easier than expected, the signal tends to be higher than the prior. The attraction of the prior pulls reported confidence below the high signal, hence below true performance on average since the signal is drawn from an unbiased distribution. While rational-Bayesian models like [Moore and Healy \(2008\)](#) may account for learning over experience, they fail to predict limited discrimination, miscalibration of confidence before completion of the task, or the absence of global learning. Therefore, we add to the Bayesian model a crucial but hidden aspect of behavior under risk or uncertainty, that is doubt. We describe the behavior of subjects who are uncertain of their true probability of success

and become consequently vulnerable to prediction errors and cognitive illusions if they rely essentially on what they perceive sequentially. We designate these subjects as "intuitive Bayesians". It turns out, unexpectedly, that the same model also predicts the overprecision bias of confidence, which we consider as a further confirmation of its validity.

Intuitive Bayesians may miscalibrate their own probability of success even if they have an unbiased estimate of their own ability to succeed. This can occur if they are uncertain of the true probability of success because they can be misled by "available" illusory signals triggered by their doubt. The direction of doubt is entirely different depending on whether their prior estimate led them to believe that they would fail or that they would succeed. We thus distinguish miscalibration among those individuals who should normally believe that they should not perform the task and those who should normally believe that they should.

To facilitate intuition, let us first consider a subject who is almost sure to succeed a task, either because the task is easy or because the subject has high-ability (H). However, the "availability" of a possible failure acts like a negative signal which leads to overweighting this possibility (Tversky and Kahneman, 1973), and underweighting her subjective probability of success¹⁶ Ep_H , *i.e.* underconfidence:

$$q_H = \mu Ep_H + (1 - \mu)0 = \mu Ep_H \leq Ep_H, \quad (1.1)$$

$$\text{with } 0 < \mu \leq 1$$

Even though high-ability agents are almost sure of succeeding the training level, their confidence is way below 1, confirming the Dunning-Kruger effect where high-

16. The time $t = (1, 5, 10)$ when confidence is reported is omitted in this sub-section to alleviate notations.

ability subjects underestimate their abilities. An estimate of this undercalibration bias for an easy task is derived from Figure 1.3a:

$$\mu_{H(\text{training level})} = \frac{0.79}{0.98} = 0.806 \cong q_{H(\text{training level})}$$

The undercalibration bias is: $1 - 0.806 = 0.194$.

However, underweighting a high probability of success need not reverse the intention of doubling. Indeed, taking the expected value as the decision criterion, among 167 "able" subjects who should double if objective probabilities are used for computation, 158 (i.e. 94.6%) still intended to double according to the subjective confidence reported before the game¹⁷.

At the other end of the spectrum, consider now a subject who is almost sure of failing, either because the task is very difficult or because the subject has low-ability (L). However, the "availability" of a possible success leads to overweighting her subjective probability of success Ep_L i.e. overconfidence:

$$q_L = \mu Ep_L + (1 - \mu)1 \geq Ep_L, \quad (1.2)$$

$$\text{with } 0 < \mu \leq 1$$

Thus, even though low ability agents should give up a difficult task, they are overconfident and are thus tempted by the returns to success¹⁸. In the limit, confidence remains positive if one is almost certain to fail. This means that low-ability individuals always exhibit a positive bottom confidence, which is in line with the Dunning-Kruger effect (they overestimate their abilities). An estimate of this overcalibration bias for the high level is derived from Figure 1.3c:

17. Very close numbers are obtained for all calibration biases with confidence reported during the game.

18. This should not be confounded with motivated inference as it applies symmetrically to undesirable and desirable outcomes.

$$1 - \mu_{L(\text{high level})} = \frac{0.34 - 0.01}{1 - 0.01} = 0.333 \cong q_{L(\text{high level})}$$

Similarly, the overcalibration bias for the middle level is derived from Figure 1.3b:

$$1 - \mu'_{L(\text{middle level})} = \frac{0.45 - 0.04}{1 - 0.04} = 0.427 \cong q_{L(\text{middle level})}$$

Notice that the overcalibration bias is about twice as large as the undercalibration bias. Hence, taking the expected value as the decision criterion, among 190 "unable" subjects who should quit if objective probabilities are used for computation, 159 (i.e. 83.7%) intended to double according to the subjective confidence reported before the game.

To sum up, we explain both the hard-easy effect and the ability effect by an availability bias triggered by the doubt about one's possibility to fail a relatively easy task (underconfidence) or to succeed a relatively difficult task (overconfidence). If probabilities are updated in a Bayesian fashion, the calibration bias is the relative precision of the illusory signal. The latter is inversely related with the absolute precision of the prior estimate and positively related with the absolute precision of the illusory signal. Thus, we mustn't be surprised to find that our estimate of the calibration bias is lower for the training level (19.4%) than for upper levels (42.7% and 33.3% respectively) because experience in the first rounds of the training level must be more relevant for predicting the probability of success in the training level than in subsequent levels. And, when comparing upper levels, the illusion of success should be more credible for the near future (middle level) than for the more distant future (high level).

This explanation is also consistent with the other measures displayed by Figures 1.3a, 1.3b, 1.3c, given the fact that they aggregate overconfident subjects who should not undertake the task with underconfident subjects who should undertake it ¹⁹. If

19. The rational decision to undertake a non-trivial task of level l (with a possibility to fail and regret) is subjective. The economic criterion for making this decision rests on the comparison of the expected utilities of all options conditional on the estimated probabilities of success at the time of decision. A rational subject should refuse the task if the expected utility of continuing to level l or above is no higher than the expected utility of stopping before level l . We make use of this criterion for writing equations 1.6 and 1.7 in the next sub-section (5.1).

λ_L is the proportion who should stop and λ_H the proportion who should continue ($\lambda_L + \lambda_H \equiv 1$), the average confidence is: $\lambda_L(\mu E_{p_L} + 1 - \mu) + \lambda_H \mu E_{p_H} = \mu \overline{E_p} + (1 - \mu)\lambda_L$. Confidence is overcalibrated on average iff $\lambda_L > \overline{E_p}$ and undercalibrated iff the reverse condition holds. The apparent overcalibration of confidence for a difficult task takes less extreme values when the average measured ability of the group rises. For instance, the results displayed by Figure 1.3c are consistent with our estimate for the overcalibration bias if the proportion of successful middle-ability subjects is 12% and that of successful high-ability subjects is 25%, since these two predicted values are close to the observed frequency of success in these groups, respectively 10% and 27%.

Remarkably, this simple model of miscalibration also predicts limited discrimination. Although Wall is more difficult than Hill at the middle level, our subjects attributed on average about the same confidence level to both tasks (see table 1.1). High-ability subjects who should double at middle level in the Wall condition, and low-ability subjects who should stop before middle level in the Hill condition would both estimate their chances of success to be higher with 16 anagrams to solve with Hill than with 20 anagrams with Wall. The former would underestimate their chances according to (1.1) and the latter would overestimate them according to (1.2), but the difference between the two estimates would be the same, equal to $\mu(E_{p_{Hill}} - E_{p_{Wall}})$. Thus, if their prior estimates were unbiased, intuitive (s.t. $\mu < 1$) high and low-ability subjects would imperfectly discriminate between Hill and Wall by underestimating the difficulty gap between them. Things are even worse for middle-ability subjects who should opt for middle level under Hill and quit before middle level under Wall. According to (1.1) and (1.2), those individuals would have a low estimate ($\mu E_{p_{Hill}}$) of their pass rate under Hill and a high estimate ($E_{p_{Wall}} + 1 - \mu$) under Wall. They would then underestimate the difficulty gap more severely than high or low-ability subjects and they might even give a higher estimate under Wall than under Hill²⁰ iff

20. It is assumed here, as in Table 1.1, that the two estimates are independent.

$E_{p_{Hill}} - E_{p_{Wall}} < ((1 - \mu)/\mu)$. Therefore, our model implies limited discrimination of differences in difficulty by intuitive Bayesians when the difference is not very salient.

A further implication of Bayesian updating is that, in the subject's mind, the precision of the posterior estimate for probabilities of success, *i.e.* confidence in her estimate, is increased by reception of the illusory signal, whatever the latter may be²¹. Therefore, *our theory of confidence predicts the overprecision phenomenon* even before completion of the task. In contrast with the other distortions of confidence, underprecision will never be observed, a prediction which is corroborated by [Moore and Healy \(2008\)](#) who do not quote any study in their discussion of "underprecision". The overestimation of the precision of acquired knowledge is an additional manifestation of the hidden search undertaken by intuitive Bayesians. Our analysis of overprecision is congruent with the observation that greater overconfidence of this kind was found for tasks in which subjects considered they were more competent ([Heath and Tversky, 1991](#)).

5. Predicting confidence biases and learning

5.1. Confidence updating by intuitive Bayesians

In our experiment, confidence is reported prior to the task E_1p , after four rounds E_5p , and after nine rounds (only for doublers) $E_{10}p$.

After going through four rounds of anagrams, a number of cues on the task have been received and processed. Participants may recall how many anagrams they solved in each round and in the aggregate, whether they would have passed the test in each round or on the whole at this stage of the task, whether their performance improved or declined from one round to the next, how fast they could solve anagrams, and so forth. For the purpose of decision-making, cues are converted into a discrete set of *i.i.d.*

21. If ν_i denotes the prior precision of subject I 's estimate of her future success (omitting level l for simplicity) $\nu_i + 1 \equiv \Phi_i$ will be the posterior precision after reception of an *i.i.d.* signal. Thus, $\Phi_i > \nu_i$. Notice that $\mu_i = \frac{\nu_i}{\nu_{i+1}}$.

Bernoulli variables taking value 1 if they signal to the individual that she should reach her goal for level l ($l = 1, 2, 3$), and 0 otherwise. The single parameter of the Bernoulli variable is its mean which defines the expected likelihood of success. However, this mean is essentially unknown to that individual. Thus, let it be denoted by \tilde{p} which is randomly distributed within the interval $[0, 1]$. Assume that the prior distribution of \tilde{p} is a Beta-distribution with a reported mean E_1p and precision ν .

Behaving like intuitive Bayesians, participants update their prior expectation of success at level l ($l = 1, 2, 3$) before the training session E_1p_l in the following manner (see DeGroot 1970, Chapter 9):

$$E_5p_l = \frac{\nu_l}{\nu_l + \tau_{4l}} E_1p_l + \frac{1}{\nu_l + \tau_{4l}} X_{1-4;l} \quad (1.3)$$

with $\tau_{4l} > 0$ designating the precision of all the independent cues perceived during the first four rounds, and $X_{1-4;l}$ defining the number of independent cues predicting future success at level l at this stage of the task. They also update the precision of the posterior expectation E_5p_l , which rises from ν_{1l} to:

$$\nu_{5l} = \nu_{1l} + \tau_{4l} \quad (3')$$

with $0 \leq X_{1-4;l} \leq \tau_{4l}$.

Equation (1.3) cannot be directly estimated on the data because the estimated probabilities E_1p_l and E_5p_l are unobservable. However, it may be rewritten concisely in terms of reported confidence $q_1(l)$ and $q_5(l)$ with the help of the miscalibration equations (1.1) and (1.2). Let us express generally the Bayesian transformation of the probability estimates into confidence as:

$$q_5(l) = \mu_{5l}E_5p_l + (1 - \mu_{5l})D_{5,l}, \quad l = (1, 2, 3) \quad (1.4)$$

with $\mu_{5l} = \frac{\nu_l + \tau_{4l}}{\nu_l + \tau_{4l} + 1}$ and ²²

$$D_{(5,l)} = \begin{cases} 1 & \text{if } \max EU(l' | E_5p_{l'}, l' = (0, \dots, l-1)) \geq \max EU(l'' | E_5p_{l''}, l'' = (l, \dots, 3)) \\ 0 & \text{otherwise.} \end{cases}$$

Confidence is merely a weighted average of the prior forecast and a *doubt term* acting as a contrarian Bernoulli signal.

And likewise:

$$q_1(l) = \mu_{1l}E_1p_l + (1 - \mu_{1l})D_{1,l} \quad (1.5)$$

with $\mu_{1l} = \frac{\nu_l}{\nu_l + 1}$ and

$$D_{(1,l)} = \begin{cases} 1 & \text{if } \max EU(l' | E_1p_{l'}, l' = (0, \dots, l-1)) \geq \max EU(l'' | E_1p_{l''}, l'' = (l, \dots, 3)) \\ 0 & \text{otherwise.} \end{cases}$$

Combining (1.3), (1.4) and (1.5), we get:

$$q_5(l) = \frac{\nu_l + 1}{\nu_l + \tau_{4l} + 1}q_1(l) + \frac{1}{\nu_l + \tau_{4l} + 1}X_{1-4,l} + \frac{1}{\nu_l + \tau_{4l} + 1}(D_{5,l} - D_{1,l}) \quad (1.6)$$

By the same reasoning, we can express the confidence of doublers for upper levels $l = (2, 3)$ as:

22. In order to have an unambiguous definition of $D_{(5,l)}$ and $D_{(1,l)}$ below, we use the expected utility (EU) criterion, as explained in note 19.

$$q_{10}(l) = \frac{\nu_l + \tau_{4l} + 1}{\nu_l + \tau_{9l} + 1} q_5(l) + \frac{1}{\nu_l + \tau_{9l} + 1} X_{5-9,l} + \frac{1}{\nu_l + \tau_{9l} + 1} (D_{9,l} - D_{5,l}) \quad (1.7)$$

with $\tau_{9l} \geq \tau_{4l}$ designating the precision of all of the independent cues perceived during the training level (9 rounds), $\nu_l + \tau_{9l}$ the precision of the posterior expectation $E_{10}p_l$, and X defining the number of independent cues predicting future success at level l at this stage of the task.

Equations (1.6) and (1.7) are essentially the same with a moving prior of increasing precision. In the absence of miscalibration, confidence reported before round t ($t = (5, 10)$) would be a weighted average of prior confidence and the mean frequency of cues predicting future success at level l since the last time confidence was reported. With miscalibration, another term is added which can only take three values, reflecting the occurrence and direction of change in subjects' estimated ability with experience. If experience confirms the prior intention to stop or continue to level l , this additional term takes value 0 and confidence is predicted by the rational-Bayesian model (with perfect calibration). However, if experience disconfirms the prior intention to stop or continue to level l , confidence rises above this reference value with disappointing experience and declines symmetrically below this reference value with encouraging experience. *Thus, our model predicts that intuitive Bayesians be conservative* and under-react symmetrically to negative experience (by diminishing their confidence less than they should) and to positive experience (by raising their confidence less than they should). Below, we report indeed rather small variations of confidence in our experiment in the form of local, but not global, learning.

5.2. Regression analysis

The models of Bayesian estimation of confidence described by equations (1.6) and (1.7) are tested by an OLS in Tables 1.4 and 1.5²³ respectively. Reported confidence in participant i 's ability to reach one level of the double-or-quits game is regressed in Table 1.4 (Table 1.5) on the confidence that she reported before the first (fifth) round and on a vector Z_{li} of level-specific cues observable in the first four (last five) rounds, assuming that $X_{1-4,li}(X_{5-9,li}) = \beta_l Z_{li} + \epsilon_{li}$ where β_l is a vector of coefficients and ϵ_{li} an error term of zero mean. Two dummy variables for the hill and choice treatments (wall as reference) have been added to the regression.

Table 1.4 – OLS estimation of the Bayesian model of confidence before round 5

	Training Level	Middle Level	High Level
Confidence before training session	0.79***	0.86***	0.90***
Freq. of rounds with 4 anagrams solved	0.14***	0.06 ^{ns}	0.01 ^{ns}
Freq. of rounds with 5- 6 anagrams solved	0.29***	0.19***	0.13***
Freq. of rounds with non-declining performance	0.12***	0.10***	0.09***
Anagrams solved per minute on rounds 1-4	0.01***	0.01***	0.01***
Hill	0.03*	0.04**	0.03*
Choice	0.01 ^{ns}	0.02 ^{ns}	0.00 ^{ns}
Constant	-0.25***	-0.25***	-0.20***
R^2	67%	70%	76%
Observations	410	410	410

Notes. **Significance level:** * $p < .10$, ** $p < .05$, *** $p < .01$, ns: not significant at 10% level. **Variables:** Frequency of rounds with non-declining performance represents the percentage of rounds (in rounds 2-4) in which number of anagrams solved was equal or higher than in the previous round, it takes four values (0,.33,.67,1). Hill and Choice: dummy variables with Wall as reference.

The regressions confirm the existence of local learning. Subjects did revise their expectations with experience of the task as several cues have highly significant coefficients (at 1% level) with the right sign. Moreover, they analyze their own performance correctly by setting stronger pre-requisites for themselves when the task gets more difficult. For example, their ability to solve just four anagrams per round in the training period increases their confidence for this period only because, if such performance is

23. The discrete value of confidence between 0 and 100 can be safely treated as continuous.

Table 1.5 – OLS estimation of the Bayesian model of confidence for doublers reported before the middle level

	Middle Level	High Level
Confidence after round 4	0.772***	0.872***
Freq. of rounds with 4 anagrams solved (5-9)	0.017 ^{ns}	-0.024 ^{ns}
Freq. of rounds with 5- 6 anagrams solved (5-9)	0.120***	0.073*
Freq. of rounds with non-declining performance (5-9)	0.034 ^{ns}	0.088**
Number of rounds used to solve 36 anagrams	0.027***	0.021**
Anagrams solved per minute on rounds 5-9	0.003 ^{ns}	-0.003 ^{ns}
Hill	-0.047***	-0.022 ^{ns}
Choice	-0.017 ^{ns}	0.022 ^{ns}
Constant	-0.136 ^{ns}	-0.186**
R^2	74%	81%
Observations	275	275

Notes. **Significance level:** * $p < .10$, ** $p < .05$, *** $p < .01$, ns: not significant at 10% level. **Variables:** Frequency of rounds with non-declining performance represents the percentage of rounds (in rounds 5-9) in which number of anagrams solved was equal or higher than in the previous round. Hill and Choice: dummy variables with Wall as reference. Number of rounds used to solve 36 anagrams (between rounds 6 and 9). (5-9) refers to measures between rounds 5 and 9.

enough to ensure success in this period, it is no longer sufficient when the task becomes more difficult. Another interesting result in Table 5 consistent with the miscalibration term in equation (1.7) concerns low achievers who double. The later they ended up solving the required number of anagrams in the training period, the more abruptly their confidence rose. It is indeed an implication of subjects' vulnerability to illusory signals that low-ability doublers find themselves almost as confident as high-ability doublers in spite of widely different performances. This result appears too on Figures 1.4 and 1.5, where the ability-adjusted confidence of low-ability doublers jumps from bottom to top during the second stage of the training period.

A major testable implication of the Bayesian model lies in the coefficient of the prior confidence, which must be interpreted as the precision of prior information relative to the information collected by experience of the task during the training period. This coefficient is always high in Tables 1.4 and 1.5 with a minimum value of 0.77. Observing such high weights for the prior favors the hypothesis of rational-Bayesian updating over adaptive expectations as the latter would considerably underweight the

prior relative to the evidence accumulated in the first four rounds. Successful experience of the easier task in the early rounds is expected to be more predictive of final success on the same task than in future tasks of greater difficulty. Thus, the relative weight of experience should diminish in the confidence equation at increasing levels or, equivalently, the relative weight of prior confidence should rise. Indeed, the coefficient of prior confidence increases continuously with the level. It rises from 0.79 to 0.86 and 0.90 in Table 1.4; and, from 0.77 to 0.87 in Table 1.5. In parallel, the coefficients of cues signaling a successful experience continuously diminish when the level rises. We can use the mathematical expressions of the two coefficients of prior confidence derived from equations (1.6) and (1.7) to calculate the precision of early experience relative to prior confidence (before the task) $\frac{\tau_{4l}}{\nu_l}$ ($l = 1, 2, 3$). With the data of Table 1.4, we get 0.266 for the training level, 0.163 for the middle level, and 0.111 for the high level. Similarly, we compute the precision of late experience relative to prior confidence (before the task) $\frac{\tau_{9l}}{\nu_l}$ ($l = 2, 3$). With the data of Table 1.5, we get 0.506 for middle level and 0.274 for high level. The impact of learning from experience appears to be substantial and with increasing returns. By elimination of ν_l , we finally calculate the precision of early experience relative to total experience during the training period $\frac{\tau_{4l}}{\tau_{9l}}$ ($l = 2, 3$). We obtain 0.322 for middle level and 0.405 for high level. The rate of increase of precision resulting from longer experience (from 4 to 9 rounds) $\frac{\tau_{9l} - \tau_{4l}}{\tau_{4l}}$ reaches a considerable 211% at middle level and 147% at high level, which forms indirect evidence of the overprecision phenomenon.

5.3. Why do intuitive Bayesians make wrong (and costly) predictions of performance?

The answer to this important question, and to the related *planning fallacy*, is contained in Table 1.6, which uses the same set of potential predictors to forecast confi-

dence in succeeding the middle level after doubling and ex post chances of success²⁴; prior confidence, ability, and performance cues observed subsequently (during rounds 5 to 9). The mere comparison of coefficients between the two columns of Table 1.6 demonstrates that posterior confidence is based on both objective performance cues and subjective variables, whereas the chances of success are predicted by the objective performance cues and ability only. The latter are the frequencies of rounds with 4 and with 5-6 anagrams solved respectively (effort) and the speed of anagram resolution (ability); and the subjective variables are essentially the prior confidence and the illusory signal given to low achievers by their (lucky) initial success. Remarkably, the number of rounds needed for solving 36 anagrams (varying from 6 to 9), which indicates low achievement and recommends quitting the game at an early stage, acts as an illusory signal with a significantly positive effect on confidence in column 1; but the same variable acts as a correlate of low ability in column 2 with a strong negative effect on the chances of success at middle level. Indeed, the subjective predictors of posterior confidence do not predict success when the objective performance cues are held constant. Prior confidence predicts the posterior confidence that conditions the decision to double²⁵ but fails to predict success because it is based on an intuitive reasoning which suffers from systematic biases. Past errors convey to the prior through the aggregation procedure of Bayesian calculus and may add up with further errors caused by the perception of illusory signals.

To reinforce our demonstration, we used the regressions listed in Table 1.6 to predict normative (based on rational expectations) and subjective (confidence-based) expected values²⁶ and determine the best choice of doubling or quitting prescribed by those al-

24. We used an OLS to predict probabilities of success so as to make the comparison with confidence transparent. Estimating an OLS instead of a Probit in columns 3 and 4 didn't affect the qualitative conclusions.

25. Conditional on initial success, prior confidence is a good predictor of the future decision to double (regression not shown). This is good news for the quality of confidence reports; and it confirms that subjects behave as intuitive Bayesians who rely on their own subjective estimates of success to make the choice of doubling.

26. The predicted values were computed on regressions containing only the significant variables. We

Table 1.6 – Estimation of posterior confidence (after doubling) and ex post chances of success at the middle level

	Level 2	
	Confidence After	Chances of success
Confidence after round 4	0.778***	0.034 ^{ns}
Freq. of rounds with 4 anagrams solved (5-9)	0.014 ^{ns}	0.276*
Freq. of rounds with 5- 6 anagrams solved (5-9)	0.107***	0.348**
Freq. of rounds with non-declining performance (5-9)	0.043 ^{ns}	-0.036 ^{ns}
Number of rounds used to solve 36 anagrams	0.024***	-0.115***
Anagrams solved per minute on rounds 5-9	0.009 ^{ns}	0.070***
Ability	-0.007 ^{ns}	0.062***
Hill	-0.046***	0.097 ^{ns}
Choice	-0.018 ^{ns}	-0.100*
Constant	-0.106 ^{ns}	0.598*
R^2	74%	30%
Observations	275	275

Notes. **Sample:** to be comparable, these regressions consider only those who succeeded first level and decided to double to second level. **Significance level:** * $p < .10$, ** $p < .05$, *** $p < .01$, ns: not significant at 10% level. **Variables:** Frequency of rounds with non-declining performance represents the percentage of rounds (in rounds 5-9) in which number of anagrams solved was equal or higher than in the previous round. Hill and Choice: dummy variables with Wall as reference. Number of rounds used to solve 36 anagrams (between rounds 6 and 9). (5-9) refers to measures between rounds 5 and 9. Number of rounds used to solve 36 anagrams (between rounds 6 and 9).

ternative models. As expected, the normative model's predictions (based on the true -ex post- probabilities) deviate farther from reality than the subjective model's: 48% versus 17% of the time. However, the confidence-based prescriptions have no information value since the rate of failure is the same whether one follows the prescription (69%) or not (70%). By contrast, the normative prescriptions have great value since the rate of failure is 52% for those who respect them versus 88% for those who don't. Finally, Table 1.7 divides the sample of doublers in four categories: 47% are able and calibrated, 12% are unable and calibrated, 36% are overconfident and 5% are underconfident. Rates of failure are markedly different among these categories: 52% only for the able calibrated, 57% for the (able) underconfident, 78% for the unable calibrated and 91% for the (unable) overconfident! Undeniably, the prevalence of miscalibration among doublers is substantial and its cost in terms of failure is massive.

checked that these values stayed close to predictions derived from the regressions listed in Table 1.6 which contain non significant variables too.

Table 1.7 – The prevalence and cost of miscalibration among doublers

Prescription of subjective expected value	Prescription of normative expected value	Category	Share	Rate of failure
double	double	able and calibrated	47%	52%
stop	stop	unable and calibrated	12%	78%
double	stop	overconfident	36%	91%
stop	double	underconfident	5%	57%

6. Conclusion

We designed an experiment analog to the popular double-or-quits game to compare the speed of learning one’s ability to perform a task in isolation with the speed of rising confidence as the task gets increasingly difficult. In simple words, we found that people on average learn to be overconfident faster than they learn their true ability. We present a new intuitive-Bayesian model of confidence which integrates confidence biases and learning. The distinctive feature of our model of self-confidence is that it rests solely on a Bayesian representation of the cognitive process: intuitive people predict their own probability of performing a task on the basis of cues and contrarian illusory signals related to the task that they perceive sequentially. Confidence biases arise in our opinion, not from an irrationality of the treatment of information, but from the poor quality and subjectivity of the information being treated. For instance, we rule out self-attribution biases, motivated cognition, self-image concerns and manipulation of beliefs but we describe people as being fundamentally uncertain of their future performance and taking all the information they can get with limited discrimination, including cognitive illusions. Above all, a persistent doubt about their true ability is responsible for their perception of contrarian illusory signals that make them believe, either in their possible failure if they should succeed or in their possible success if they should fail.

Our intuitive-Bayesian theory of estimation combines parsimoniously the cognitive

bias and the learning approach. It brings a novel interpretation of the cognitive bias and it provides a general account of estimation biases. Indeed, we did not attribute confidence biases to specific cognitive errors but to the fundamental uncertainty about one's true ability; and we predicted phenomena beyond the hard-easy and Dunning-Kruger effect which could not be explained all together by previous models: miscalibration and overprecision *before* completion of the task, limited discrimination, conservatism, slow learning and planning fallacy. Moreover, we showed that these biases are likely to persist since the Bayesian aggregation of past information consolidates the accumulation of errors, and the perception of illusory signals generates conservatism and under-reaction to events. Taken together, these two features may explain why intuitive Bayesians make systematically wrong and costly predictions of their own performance. Don't we systematically underestimate the time needed to perform a new (difficult) task and never seem to learn?

Our analysis of overconfidence is restricted to the overestimation bias. The latter must be carefully distinguished from the overplacement bias since the hard-easy effect that we observed here with absolute confidence has often been reversed when observing relative confidence: overplacement for an easy task (like driving one's car) and underplacement for a novel or difficult task. The reasons for overplacement are probably not unique and context-dependent. When people really compete, the over (under) placement bias may result from their observing and knowing their own ability (although imperfectly) better than others'. If both high-ability and low-ability individuals compare themselves with average-ability others, the former are likely to experience overplacement and the latter underplacement. The same reasoning applies to individuals familiar or unfamiliar with the task, and to individuals who were initially successful or unsuccessful with the task. When no real competition is involved, the overplacement effect relates to an evaluation-based estimate of probability. While there is an underlying choice to be made in the estimation task, no such thing is present in the latter

case. If I ask you whether you consider yourself as a top driver (relative to others), I don't generally expect you to show me how you drive. Preference reversals are not uncommon between choices and evaluations (Lichtenstein and Slovic, 1971). Thus, the present analysis of overestimation is consistent with reasonable explanations of overplacement. Moreover, it predicts the overprecision phenomenon and even rules out underprecision. This demonstrates that overestimation and overprecision are related but different biases.

Double-or-quits-type behavior can be found in many important decisions like addictive gambling (Goodie, 2005), military conquests (Johnson, 2004), business expansion (Malmendier and Tate, 2005), speculative behavior (Shiller, 2000), educational choices (Breen, 2001), etc. Overconfident players, chiefs, entrepreneurs, traders, or students are inclined to take excessive risks; they are unable to stop at the right time and eventually fail more than well-calibrated persons²⁷ (e.g., Barber and Odean 2001, Camerer and Lovallo 1999). In contrast, under-confident individuals won't take enough risks and stay permanently out of successful endeavors.

On the theoretical side, the intuitive-Bayesian model of confidence before completion of a task creates a link between confidence and decision analyses and their respective biases. Confidence biases and the anomalies of decision under risk or uncertainty can be analyzed with the same tools. The estimation of one's ability implies an implicit comparison between an uncertain binary lottery and a reference outcome. It is a by-product of the question: should I double or quit? This is a question of interest to behavioral and decision theorists.

27. However, overconfidence may pay off when there is uncertainty about opponents' real strengths, and when the benefits of the prize at stake are sufficiently larger than the costs (e.g., Johnson and Fowler 2011, Anderson et al. 2012).

A. Appendices

Figure 1.7 – Example of the task screen

The screenshot shows the 'ETAPE 5' task screen. On the left, a list of anagrams is provided: 'eeenps', 'imate', 'itopn', 'hiecnm', 'reafi', and 'ormbe'. Next to each is an input field. The 'itopn' field contains 'point' and has 'OK' feedback. The 'hiecnm' field contains 'cheimn' and also has 'OK' feedback. To the right, a summary box shows: 'Nombre de mots trouvés: 2', 'Nombre de mots à trouver: 6', 'Bilan étapes 1 à 9: 15', and 'Nombre minimal à réaliser: 36'. Below this is a 'PASSER' button. At the bottom, a timer labeled 'Temps Restant (mins)' shows 3, 2, and 1 minutes remaining.

Legend:

- A: Actual round (round 5 in this example).
- B: List of anagrams to be decoded.
- C: Fields to type the correct word.
- D: Feedback. The "OK" appears when the solution for the anagram is correct.
- E: Number of correct anagrams in the current round.
- F: Total anagrams to be decoded in the current round, 6 in this example (first level).
- G: Number of cumulated correct anagrams, including the current and previous rounds.
- H: Number of correct anagrams required to solve the current level, in this example 36 (first level).
- I: Remaining time. The total time is 8 minutes, we show only the 3 last minutes.
- J: Button to go to next round. Participants can pass to next round without clearing all anagrams in the current level, but they cannot come back once they pushed the button.

CHAPTER 2

CREATIVE POTENTIAL AND ECONOMIC
BEHAVIOR

1. Introduction

Creativity has been defined as “the ability to produce work that is both novel and appropriate” (Sternberg and Lubart, 1996). Translated into economic language, creativity is “the ability to produce efficient innovations¹” and should be an important economic concern. However, few economic studies have dealt with creativity or creative behavior so far. The scarcity of research on creative behavior in the economics field can be attributed to two main factors: (i) economists neglected the impact of non-cognitive abilities² on economic outcomes up to the last decade (Borghans et al., 2008); (ii) the assessment of creativity potential is still a bottleneck in the economics research, these measures are not yet applicable in large scales, being restricted to experimental studies. The existing economic literature lies in the fact that production and consumption of new products are uncertain activities, implying risk taking and entrepreneurial skills associated with the creative behavior (Menger and Rendall, 2014). The first goal of this paper is to review how economists describe creative behavior and propose how it should be described. The second goal is to observe how the potential of creativity impacts individuals’ production. For that, we introduce several measures of creativity in economic experiments with adolescents, and additionally we evaluate the impact of creativity on school achievements -a "real life" performance.

1.1. Creativity in economic literature

As a first step, an economist would ask whether creative workers are more productive than others, that is, whether creativity enhances their human capital or market skills. In the same way that investments in R&D, education, training, and the quality of labor

1. Innovation refers to the implementation of something new, such as products, services or processes of production.

2. In economic literature, cognitive ability refers to mental skills that are related to knowledge (memory, reasoning, learning, decision making); other skills such as personality traits and motivation are called *non-cognitive*. Creativity potential conveys both cognitive and non-cognitive skills.

were initially introduced into economic analysis as the “residual factors” of economic growth when the latter was explained essentially by quantities of labor and physical capital (Denison et al., 1962), can creativity and other personality variables explain the large residual of human capital earnings functions³? This question has attracted a lot of attention among labor economists in the last fifteen years. Bowles, Gintis and Osborne’s (2001) influential survey of the determinants of earnings rate “non-cognitive personality variables, such as attitudes towards risk, ability to adapt to new economic conditions, hard work, and the rate of time preference” as potential factors of labor market success. Although not explicitly cited here, creativity was added to this list by other researchers, such as Sternberg (2000, 2001) and Garcia (2014). Borghans et al. (2008) set an important bridge between differential psychology and economics, confirming the predictive power of non-cognitive abilities and suggesting that they may play a role in many economic outcomes. The main idea is that each individual is endowed with a set of innate abilities and skills that are responsible for future outcomes, including schooling (García, 2013).

The assessment of the psychological skills is a challenge in economic applications. Some measures, such as the Big Five (Costa and MacCrae, 1992), have been simplified and can now be included in household surveys⁴. Other measures, including creativity, must still be assessed in the experimental laboratory⁵.

3. Human capital earnings functions predict the logarithm of earnings with years of education, and training, and a quadratic function of market experience (Mincer, 1974). When available, a measure of cognitive ability like IQ is added.

4. For instance, the BHPS (British Household Panel Survey) has included a brief questionnaire to assess the personality traits in some waves.

5. There exist web-based platforms designed to assess the individual potential of creativity, such as the Creative Profiler developed by Paris Descartes University’s psychologists’ team (<http://www.creativeprofiler.com/>). This kind of tool can reach and assess creativity more broadly than in the lab, but there are still some issues concerning research in economics: it takes more than one hour to be responded, scores must be evaluated by a jury, and it is still hard to link the individual creative potential to existing economic outcomes.

1.2. An economic approach to creativity

Economics and psychology have different approaches to creativity because they have different perspectives. Whereas psychology is concerned with the decision process, economics has something to say about the innovation process. We begin by contrasting these two approaches before looking for their complementarity and convergence.

Psychologists distinguish two aspects that are fundamental in the production of creative work: the divergent-exploratory and the convergent-integrative processes (Barbot, Besançon and Lubart, 2011). Table 2.1 provides the definition and the main individual characteristics requested by them.

Table 2.1 – Aspects in the production of creative work (Barbot, Besançon and Lubart, 2011)

	Divergent-exploratory	Convergent-integrative
Definition	generating new ideas and solutions for existing problems	synthesis of disparate existing ideas into a novel and appropriate solution
Subprocesses	flexibility, divergent thinking, and selective encoding	associative thinking, and selective comparison
Individual characteristics	openness to experiences and intrinsic task-oriented motivation	tolerance for ambiguity, perseverance, risk taking, and achievement motivation

Thus, the creative decision process is decomposed into a phase of mental divergence followed by a phase of mental convergence, working in cycles. Mental divergence allows finding new ideas to problems; while mental convergence allows the synthesis of disparate ideas into a novel and appropriate solution. Obviously, both traits are useful for innovating and must act in coordination because new ideas don't fall from heaven, they come to the mind. Many things come to the mind all the time though, but, if the person is focused in a specific direction, she might lend attention to a signal and convert it into a valuable idea if she is endowed with a good sense of serendipity. What seemed to indicate mental divergence, i.e. finding new ideas, requires a form of mental convergence. Things that come to the mind are not automatically interpreted as signals: most of them will be probably dismissed as noise and forgotten, and only appropriate target-directed ideas will be recalled. Thus, divergence and convergence

are complementary within the creative personality. Intelligence is another important aspect of the creative personality, according to [Karwowski et al. \(2016\)](#) it is a necessary (but not sufficient) condition of the creative potential. Thus, we must expect a positive correlation when investigating the impact of creativity on school achievements.

From an economic perspective, creative behavior must be judged by an individual's propensity to innovate in production and consumption activities. Creative behavior in production implies in finding new ideas and techniques that will raise productivity and output. A creative person is not expected to be more productive than others in repetitive well-known tasks, but the divergent-exploratory aspect of creativity suggest that she should be more inventive by searching more than others and finding (new) solutions to unresolved issues or better solutions to old problems. Search may be a valuable activity because we have limited knowledge of the existing world and/or because the world changes unexpectedly. The ability to search is advantageous in a number of circumstances and disadvantageous in others. Perseverant search and exploration of alternatives enhance the rate of discovery of beneficial innovations in either production or consumption. In the context of production, such behavior prolongs the range of increasing marginal returns which are responsible for endogenous economic growth ([Romer, 1986](#)). If, on one hand innovation can be associated to economic development, on the other hand the lack of development of new products can cause recession and financial crisis ([Hausman and Johnston, 2014](#)). In the context of consumption, random search allows creative consumers to reach their true preferences and thus causes a permanent rise in their utility ([Armantier et al., 2015](#)). However, repeated search forces the explorer to forego her currently best choice for some time and the short run costs must be balanced with the long run benefits of innovation. The research ability described here should correlate with mental flexibility, divergent thinking and a sense to select relevant information. In a changing environment, a new type of search emerges: the entrepreneurial capacity to adapt to unexpected changes and reap unexpected op-

portunities (Schumpeter, 1934). In contrast with the research ability, entrepreneurship is a capacity which is not a “skill”, that is, an ability to produce work efficiently with existing technology, and it does not raise job performance in most jobs. Indeed, creative workers are ranked poorly by supervisors within work groups (Edwards, 1977). An entrepreneurial mind is useless or even harmful in a static, unchanging, world. It only finds its place and justification in a changing, dynamic, world. Schultz (1975) attributed a significant portion of the economic return to education to this capacity which he designated as “the ability to deal with disequilibria” because the rents captured by entrepreneurs after technological shocks only survive in a competitive economy until markets reach a new equilibrium.

What the previous discussion has shown is that we must distinguish two types of economic innovators: researchers (the ability to find new solutions) and entrepreneurs (the ability to capture unexpected rents). These two types of creative behavior are equally found in production and consumption, notwithstanding this study focus on production. How do the creativity scores derived from psychological research match with this economic typology?

From the foregoing discussion, we derive three testable hypotheses:

- **H1.** Creative individuals are no more productive than others in repetitive tasks;
- **H2.** Creative individuals of the researcher type are more productive than others in research activities;
- **H3.** Creative individuals of the entrepreneurial type are more able than others to capture unexpected rents.

This paper is organized as follows: section 2 describes the experimental design and presents descriptive statistics. Hypotheses are tested in section 3, while section 4 evaluates the impact of creativity on school achievements. Conclusions are presented in section 5.

2. The experiment

2.1. Task and treatments

These hypotheses are tested below on experimental data collected in the “*Laboratoire d’Economie Expérimentale de Paris*” in December 2014 on 169 high school students from the Paris region (divided in three school districts in Paris, Créteil and Versailles). The students were 10th graders attending a one-day conference cycle about economics including participation to a laboratory experiment. 76.3% of the students are aged 14 or 15, 23.7% are aged 16 or 17, with a fairly equal number of boys and girls. One justification for observing adolescents is that we can disentangle the respective effects of educational attainment and creativity by observing both simultaneously on high school students of the same class. This young population of similar individuals controls for a number of common factors and is immune from other unobservable factors that might have appeared later in life and caused additional differentiation between the subjects. Nine experimental sessions over three days (one per school district) were conducted, each session involving about 20 participants and lasting 90 minutes. The experiment consisted of personality measures, two creativity scores, two independent real-effort tasks including comprehension and satisfaction questions, and a general questionnaire including scores obtained in various fields at the “*Brevet des collèges*” - a nationwide exam taken after the 9th grade. The incentives related to the two real effort tasks, an amount of candy or chocolate proportional to the gains have been surprisingly well received⁶. Some of the students won an entire bag of candies or chocolates, while others did not win anything.

The Five-Factor model of personality traits, also called Big Five, is widely used and lots of researches in various fields confirm its validity. The five dual dimensions derived

6. Announcing the names of the two well-known brands that produce the candies and chocolates seemed to motivate the students even more.

from Big Five tests are extraversion, agreeableness, conscientiousness, openness to experience and emotional stability (or neuroticism). Because of the experiment's length, we chose a short test, the French version of TIPI (Ten Item Personality Measure) developed by [Gosling, Rentfrow and Swann \(2003\)](#), which takes about two minutes to complete⁷, has satisfactory psychometric properties and good levels of temporal stability. Comparisons with other brief Big Five tests show that the TIPI performs well in terms of convergence and reliability, and has a valid version in French ([Storme, Tavani and Myszkowski, 2016](#)). As we observe adolescents who are beginning to form ideas about their future, we completed this popular description of personality by the vocational typology⁸ of [Holland \(1966\)](#): Realistic (Doer) , Investigative (Thinker), Artistic (Creator), Social (Helper), Enterprising (Persuader), and Conventional (Organizer)⁹. Research from [Costa, McCrae and Holland \(1984\)](#) has established strong associations between vocational preferences and personality dispositions. For example, "investigative" scores are correlated with openness and introversion.

We added to the questionnaire self-reported questions (on a 0 to 10 Likert scale) about "willingness to take risks" in general ([Dohmen et al., 2011](#)) and in the specific domains of health, leisure and sport, school, and other (being honest or lying towards friends and family)¹⁰. The rationale for introducing risk aversion is that a positive correlation of risk tolerance with creativity is expected (as explained in next section)

7. In this version, each item consists of two easy "descriptors", separated by a comma, using the common stem "I see myself as:" (for example "Sympathetic, warm" or "Conventional, uncreative"). The students have to rate each of the ten items on a 7-point Likert scale ranging from 1 ("disagree strongly") to 7 ("agree strongly"), even if one characteristic applies more strongly than the other. Each personality factor is measured in a positive and in a negative (reversed) way to "represent both poles of each dimension": extraversion corresponds to items 1 and 6 reversed, agreeableness to items 2 reversed and 7, and so on. The even items have thus to be reverse-scored (7 recoded with 1, etc.); the final measure of one factor is given by the average of the two items (the standard item and the recoded reverse-scored item).

8. In practice, the participants had to rate thirty professions on a scale ranging from 0 to 4.

9. Table 2.26 summarizes the characteristics of the six vocational types and gives examples of professions associated with them.

10. We have adapted the idea of risk in specific domains proposed by [Dohmen et al \(2011\)](#) to the public of adolescents, since we excluded the question about risk when driving, and included the risk at school (preparing homework, studying before exams, not being absent, ...)

as was already found with men and IQ (Bonin et al., 2007).

The two measures of creativity are taken from Lubart, Besançon and Barbot's EPoC battery (*Evaluation du Potentiel Créatif*, 2011), which evaluates creative potential among children at school. This procedure has a great validity, authors found a high and significant correlation with the divergent thinking and the traditional Torrance's test of creative thinking (Torrance, 1962), moreover they found a correlation between the creativity measures of the EPoC to *openness* personality trait, in line to McCrae and Costa (1987) observations that *openness for new experiences* facilitates divergent thinking.

For timing reasons, we selected two graphical sub-tests (among the whole set of eight verbal and graphical sub-tests) preceded by a verbal warm-up test¹¹. We chose tests of graphical creativity for two main reasons: first, drawing is less culturally-dependent than writing and is not a high-school-relevant skill, thus may be more fair and exogenous; second, drawing is unexpected and generally amusing even for 10th graders. In each test's instructions, participants were asked to be original and to draw something different from what the others would do. The tests were as follows:

- **Graphical divergent-exploratory thinking (DT)** from a concrete stimulus in 10 minutes: making as many drawings as possible using the object that appears on the screen in different and unusual manners. The object is a banana. Drawings were made on several A5 sheets of paper that were on the participants' desks alongside with pencils. The tests are rated by simply counting drawings and excluding the off-topic ones. The higher the number of drawings, the higher the individuals graphical divergent thinking potential. Figure 2.7 in the Appendix shows some examples of this task.
- **Graphical convergent-integrative thinking (IT)** from a concrete stimulus in 15 minutes: inventing a unique drawing using at least four out of the eight objects

11. Verbal divergent thinking in 3 minutes (warm-up): imagining and writing down on paper all the possible uses of a stick of wood. The score of this test relies in the number of outputs.

appearing on the screen. The objects are a jointed manikin, a lamp, a sunhat, a spade, a stone, a dolphin plush, a bag and a carrot. Drawings were made on one A4 sheet of paper and felt pens were additionally given. The rating of the integrative thinking test relies on a scale ranging from 1 to 7, "1" meaning "Very poor, total absence of ideas" and "7" meaning "The majority of elements used in another context, all elements integrated in an innovating way". Originality and integration of the objects were the two main rating criteria. Three correctors¹² rated the drawings based on a objective scale provided in the manual of the Epoc test and converged to a score for each drawing¹³.

Two incentivized real-effort tasks were designed to test the hypotheses about creativity and economic behavior listed above. The three *buttons tasks* aim at assessing the agents' search behavior whereas the *typing task* is purely repetitive and almost non-cognitive. The *buttons tasks* consist in three series of a hundred clicks on four or eight buttons. Each click and each button give a different random number of points that can be positive, null or negative; each button holds a (hidden) pre-fixed distribution of values, the same for all participants. The ordering of buttons is not the same among participants. One has to obtain the highest cumulated number of points¹⁴. One of the three series is randomly selected at the end of the task and its outcome determines the reward for this part. The first series doesn't require any creativity, the second series detects creativity of the researcher-type, and the third series detects creativity of the entrepreneurial-type¹⁵.

12. I would like to thanks Marie Thillot (LATI -Université Paris Descartes) and Gabrielle Tallon (Université Paris 1) for the valuable contribution in this process of rating the creativity tests.

13. Figures 2.8, 2.9 and 2.10 in the Appendix provide some examples of the task and scores.

14. Participants are randomly assigned to two different incentive conditions: performance and competition. In the "performance" treatment, the total gain only depends on the individual performance (X points correspond to Y candies). In the "competition" treatment, the total gain also depends on the performances of the other competitors; the first three receive a reward, the others do not receive anything. We consider the competition condition in our regressions, but there are no impacts over outcomes.

15. We present in the Appendix (section A.3) some statistics about the hidden outcomes for each series of the buttons task.

- First series: **"equal solution"** (four buttons on the screen). Points are distributed over the four buttons and clicks so that each button yields the same number of cumulative points each 10 clicks, and consequently after the 100 clicks. There is a different variability of points between buttons, but as they show the same accumulated values every 10 clicks, lack of exploration is not harmful for this task and may even be beneficial.
- Second series: **"best button"** (eight buttons on the screen). Points are distributed over the buttons and clicks so that after a number of trials one can eventually recognize the best button. The best button yields a higher number of accumulated points each 10 clicks, but not the higher outcome every click. The 96th click on the "best button" gives a "jackpot" reward of 200 points, intended for a participant who recognized the "best button" and stuck to it until the end. It is important to clarify that the button is the best even if the player misses the jackpot. This series has been designed to elicit a "researcher" type of work: there are lots of paths (buttons) to explore, one is definitely the best but it takes time to discover this truth and eventually win the jackpot.
- Third series: **"best moving strategy"** (four buttons on the screen). Points are distributed over the buttons and clicks so that there are three successive "best strategies": there is a best button as in the second series, but the best button changes twice during the series. This series has been designed to elicit an "entrepreneurial" type of work: one has to find the best strategy quickly, know how long to stick to it and when to move to a new strategy that will be superior in the near future. There are 3 possible "jackpots" of 100 points at the end of each "best button" series.

In the *typing task*, participants have to retype the highest possible number of codes in three minutes. The codes appear as a list on each participant's screen. A code

consists of five random letters¹⁶ that do not form a word. Once correctly retyped, the code is highlighted in the list. The typing task was performed four times - twice alone, twice with a randomly assigned partner participating in the same session whose picture appears on the screen (without communicating). To avoid biases, the ordering changes for each session. All participants performed the task twice with the "performance" treatment described previously, and twice with the "competition" treatment.

In contrast with the typing task, the buttons task requires memorizing previous gains and buttons to optimally choose on which button to click next. Thus, a significant relation between grades at school and gains is expected for the buttons tasks.

2.2. Descriptive statistics on creativity scores

We work with three scores of creativity: the graphical **divergent thinking (DT)**, the graphical **integrative thinking (IT)** and the graphical **creativity index (CI)**, that is a composition of the normalized values for the two previous measures¹⁷. Table 2.2 presents the descriptive statistics for the three creativity scores and figure 2.1 shows the distribution of these variables. The distribution of the creativity index resembles a normal distribution (*skewness* = 0.21, and *kurtosis* = 2.88). Moreover, the distribution of integrative thinking is symmetric (*skewness* = 0.00), while divergent thinking distribution of scores is right skewed (*skewness* = 0.75)¹⁸, the same pattern observed by Berlin, Tavani and Besançon (2016) with a younger sample. There are evidences, not yet confirmed, that divergent thinking is developed later than integrative thinking (Lubart, Besançon and Barbot, 2011).

16. We consider only characters that are used in French to create our codes.

17. We use normalized creativity scores in order to make them comparable and simplify analysis.

18. Means for divergent and convergent scores are statistically different ($p = 0.0001$, paired t-test)

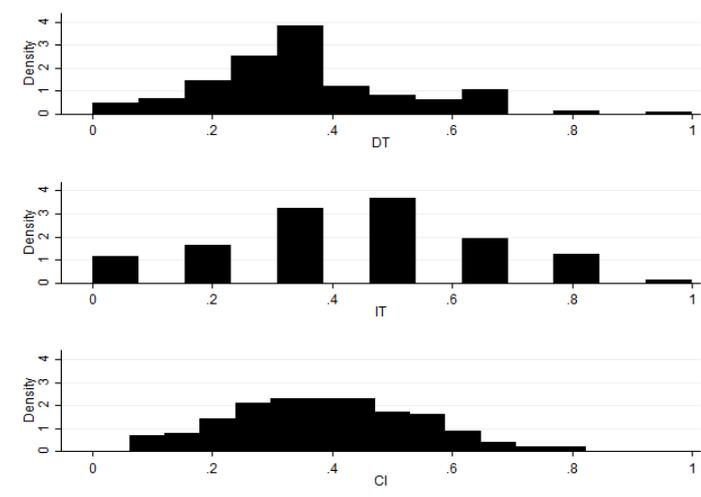
19. Skewness measures the lack of symmetry of a distribution. If the distribution is symmetric, the coefficient of skewness is 0. If the coefficient is negative, the median is usually greater than the mean and the distribution is said to be skewed left. If the coefficient is positive, the median is usually less than the mean and the distribution is said to be skewed right.

20. Kurtosis measures the peakedness of a distribution (curvature). The smaller is the coefficient of kurtosis, the flatter is the distribution. The normal distribution has a coefficient of kurtosis of 3 and

Table 2.2 – Normalized creativity scores

Creativity (scale 0-1)	Mean	Median	Std. Dev.	Skewness ¹⁹	Kurtosis ²⁰	n
DT	0.35	0.31	0.17	0.75	3.83	169
IT	0.43	0.38	0.24	0.00	2.49	169
CI	0.39	0.38	0.15	0.21	2.88	169

Figure 2.1 – Distribution of the three normalized scores



Correlations between the three creativity measures²¹ are reported in Table 2.3. They are not high (between 0.13 and 0.33) but always significant. These results are quite similar to the ones found by Berlin, Tavani and Besançon (2016)²², which is a sign of the test's robustness. The fact that IT and DT are not very correlated (0.13) also justifies our choice to consider them separately in our statistical treatments. In particular, they might have offsetting effects, neutralizing the overall score effect.

To have an overview of creative agents' characteristics, we compute the correlations between our creativity scores and demographic, personality and risk-taking variables.

provides a convenient benchmark.

21. We include the verbal divergent thinking (our warm-up task) in the correlation analysis in favor to observe the robustness of our measures. The verbal divergent thinking is not used to describe our results because adolescents had only 2 minutes to perform the warm-task, instead of the 10 minutes proposed by Epoc.

22. Berlin, Tavani and Besançon (2016) found correlation coefficients between 0.19 and 0.26 for the same measures

Table 2.3 – Creativity correlation table with p-values in parentheses (Pearson’s test)

	Verb. DT	Graph. DT	Graph. IT
Verb. DT	1.00		
Graph. DT	0.3098 (0.0000)	1.00	
Graph. IT	0.3309 (0.0000)	0.1298 (0.0926)	1.00

Table 2.4 shows that the our creativity scores are orthogonal to age, gender²³ and the two personality measures that we consider (Big Five and Vocation), with few exceptions that will be discussed in the next paragraph. Thus, the potential of creativity constitutes a new parameter of interest that does not resemble purely non-cognitive ones. The only exception shown by Table 2.4 to the alleged independence between creativity and other personal characteristics concerns risk-taking. Highly creative agents are significantly more risk-seeking in "general", and regarding "health" and "leisure". Indeed, risk-taking is one of the features of creative personalities observed in the existing literature (Glover and Sautter, 1977, Sternberg and Lubart, 1991a).

A closer look to the relation between creativity scores and personality measures shows that correlation coefficients are weak and mostly not statistically significant. The creativity index correlates only to Extraversion ($p = 0.0253$), while integrative thinking correlates to Extraversion ($p = 0.0362$) and Investigative ($p = 0.0815$). Divergent thinking is positively correlated to openness ($p = 0.0523$), this is coherent to the findings of McCrae and Costa (1987) that openness mediates unusual ideas, imagination and curiosity. Moreover, according to Lubart, Besançon and Barbot (2011), divergent thinking potential involves knowledge, search mechanisms, personality traits such as openness and perseverance. This characteristics about divergent thinkers suggest the identification of the creative person from a psychological perspective with what we qualified as a “researcher” from an economic perspective. Moreover, the psychological

23. Lubart, Besançon and Barbot (2011) come to the same conclusion, namely that gender plays a negligible role in creative potential.

Table 2.4 – Creative agents' characteristics (correlations)

Characteristics	DT		IT		CI	
	<i>r</i>	<i>sig.</i>	<i>r</i>	<i>sig.</i>	<i>r</i>	<i>sig.</i>
Women	-0.0434	<i>ns</i>	-0.0629	<i>ns</i>	-0.0722	<i>ns</i>
Age	-0.0162	<i>ns</i>	-0.1155	<i>ns</i>	-0.0982	<i>ns</i>
Personality (scale 1-7)						
Extraversion	0.0873	<i>ns</i>	0.1613	**	0.1720	**
Openness	0.1496	*	-0.0301	<i>ns</i>	0.0575	<i>ns</i>
Agreeableness	0.0401	<i>ns</i>	-0.0272	<i>ns</i>	0.0006	<i>ns</i>
Conscientiousness	0.0036	<i>ns</i>	-0.1073	<i>ns</i>	-0.0812	<i>ns</i>
Emotional stability	-0.0986	<i>ns</i>	-0.0729	<i>ns</i>	-0.1097	<i>ns</i>
Vocation (scale 0-4)						
Realistic	0.0387	<i>ns</i>	0.0449	<i>ns</i>	0.0556	<i>ns</i>
Investigative	0.0069	<i>ns</i>	0.1342	*	0.1002	<i>ns</i>
Artistic	-0.0639	<i>ns</i>	0.0025	<i>ns</i>	-0.0326	<i>ns</i>
Social	-0.1042	<i>ns</i>	-0.0328	<i>ns</i>	-0.0817	<i>ns</i>
Conventional	-0.0949	<i>ns</i>	0.0056	<i>ns</i>	-0.0470	<i>ns</i>
Enterprising	-0.1227	<i>ns</i>	0.0219	<i>ns</i>	-0.0493	<i>ns</i>
Risk taking (scale 1-10)						
Risk general	0.0896	<i>ns</i>	0.1322	*	0.1508	*
Risk health	0.1561	**	0.1282	*	0.1836	**
Risk leisure	0.1509	*	0.1258	<i>ns</i>	0.1789	**
Risk others	0.0073	<i>ns</i>	-0.0368	<i>ns</i>	-0.0246	<i>ns</i>
Risk school	0.0835	<i>ns</i>	0.0243	<i>ns</i>	0.0639	<i>ns</i>

r: Pearson's correlation coefficient. Significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

index of creativity overlooks the entrepreneurial type so prominent in the economic perspective.

3. Testing economic predictions

Performance in the buttons tasks is measured by the cumulative number of points obtained for each task since the players had an economic incentive (candies) to get the highest cumulative score. However, players clicked on buttons without knowing the pre-set number of points that would come out. Similarly, performance in the typing task is measured by the number of correct codes retyped. At the outset of this section, we predicted that: (H1) creative individuals are no more productive than others in simple repetitive tasks; (H2) creative individuals of the researcher type are more productive

than others in research activities; (H3) creative individuals of the entrepreneurial type are more able than others to capture unexpected rents. Since the psychological creativity index that we used here does not capture the entrepreneurial dimension of economic creativity as argued in the last paragraph, we expect to find no effect of the available index on performance in the "best moving strategy" series (H3 revised). Indeed, the performance in the "best moving strategy" series provides a preliminary measure of creativity of the entrepreneurial type that will be used in section 4 to assess the effects of economic creativity on school achievements. An additional test relying on the investment nature of search behavior is provided below: (H4) creative subjects are engaged in costly search when performing creative tasks, but not when performing simple repetitive tasks. Table 2.5 displays tests of H1 on the typing task. Then, Table 2.9 presents tests of H2, H3 revised. Tests of H4 are shown in Table 2.13.

3.1. Typing: a simple repetitive task

Result 1: Creative individuals are no more productive than others in simple repetitive tasks. We present in Table 2.5 columns (1) and (2) the relation between the number of correctly retyped codes and our measures of creativity, controlled by treatments, "disorder", "incorrect codes" and other variables. In this table, disorder corresponds to the number of correctly retyped codes that were not following the last one. Participants could either retype codes one after another on a row, or pick codes randomly in search of an easy one or to have fun. No technique was superior. Each correctly retyped code was highlighted so that there was no waste of time searching. The number of incorrect codes corresponds to codes typed wrongly, each code typed wrongly is a waste of time and may decrease the performance in the task. Table 2.5 corroborate H1 since there is no effect of creative potential over the task output: creative individuals are no more productive than others in this kind of task. Moreover, it is possible to observe that the number of correct codes is mainly affected by treatments (pair

and competition), number of codes typed incorrectly and risk seeking behavior, but not by other variables presented in the table. Risk-seekers should be more motivated by an incentivized experiment offering an opportunity of "gambling" ²⁴.

Table 2.5 – Typing task: outcomes and efficiency

OLS with clustered standard errors ²⁵	Number of correct codes		Efficiency (%)	
	(1)	(2)	(3)	(4)
DT	-1.9409 (4.0758)		-.0043 (.0299)	
IT	4.4303 (3.3644)		.0223 (.0207)	
CI		3.8464 (4.5047)		.0237 (.0381)
Pair	-1.2061** (.5128)	-1.2022** (.5134)	-.0105 (.0065)	-.0105 (.0065)
Competition	-.3815 (.5437)	-.3749 (.5414)	-.0079 (.0061)	-.0079 (.0061)
Pair × Competition	2.8253*** (.6810)	2.8142*** (.6817)	.0231** (.0090)	.0230** (.0090)
Incorrect codes	-.9005* (.5077)	-.9255* (.5171)		
Disorder	-.4047 (1.2861)	-.4753 (1.2995)	-.0290*** (.0075)	-.0294*** (.0075)
Risk seeking	.6371** (.2489)	.6331** (.2516)	.0038 (.0023)	.0037 (.0023)
Age	-1.1371 (1.5114)	-1.2850 (1.5264)	-.0011 (.0085)	-.0017 (.0085)
Women	.1818 (1.2937)	.0403 (1.2666)	.0110 (.0083)	.0104 (.0083)
Constant	35.8582 (21.8610)	38.4391* (21.9692)	.9797*** (.1404)	.9901*** (.1397)
R^2	.2216	.2126	.1551	.1529
Adjusted R^2	.1762	.1687	.1080	.1078
Observations	436	436	436	436
Clusters	109	109	109	109

Notes: **Standard errors** in parentheses. **Significance level:** * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. **Disorder:** typing in disorder is a dummy variable for each of the four rounds participants performed the task: 0 if participant typed all codes in order and 1 if typed at least one code in disorder. Descriptive statistics for efficiency and disorder are presented in Table 2.15. **Control variables:** academy, session, order of the sub-task, comprehension errors, and big five personality traits. **Sample size:** due to technical reasons, we have only 109 adolescents participating in the typing task.

24. Since creative people tend to be risk-seekers, including risk-seeking as a variable might crowd-out the effect of the creativity variables. Thus, we perform same regressions without "risk-seeking" and global results for creativity measures are still the same.

25. The number of correct codes is a count variable. We select an OLS regression for four reasons. First, the distribution of this variable is more similar to a Gaussian distribution than a Poisson distribution (see Figure 2.4 and Table 2.19 in the Appendix to this Chapter for more details). Second, OLS coefficients have the same direction and significance level when compared to a negative binomial regression or a poisson regression (see this comparison in Table 2.22 in the Appendix to this Chapter). Third, the OLS regression model does not produce negative predicted values. Fourth, we are interested in the analysis of several interaction terms in Tables 2.5 and 2.7. According to Ai and Norton (2003)

In the same table columns (3) and (4), efficiency corresponds to the ratio between the number of correct codes retyped and the total of codes retyped. We observe that, as the output, efficiency in the task is not impacted by our creativity measures.

When comparing predictors for output and efficiency, Table 2.5 shows that adopting the strategy of typing in disorder does not impact the performance in the task, but it is inefficient. As codes are highlighted if correctly typed, the chances of producing less mistakes are higher when typing in order since it is easier to find the code being typed if one needs to review the code before retyping it correctly.

Result 2: Creative potential produces order, but does not enhance output and efficiency in this simple repetitive task. Creative individuals picked codes more frequently in an orderly fashion than others (Table 2.6). Increasing the IT score in 0.1 point²⁶ decreases by 2.8% the chances of typing in disorder ($p = 0.090$), while an increment in 0.1 point for CI scores represents 3.8% less chances of selecting codes in disorder ($p = 0.092$). Although such behavior may look conventional at first sight, a straightforward interpretation exists: creative individuals are intelligent and intelligent individuals won't waste time searching randomly if there is no need to search. Apparently creative individuals select the better strategy, but they don't convert it in higher output and efficiency, perhaps because they are able -and prefer - to spend less effort on this repetitive task.

Result 3: IT-gifted subjects are more cooperative than others. The apparently different effects of competition in pairs in Table 2.7 are fully consistent with those described in Table 2.5 on performance and efficiency if IT-gifted subjects are assumed to be more cooperative.

"the magnitude of the interaction effect in nonlinear models does not equal the marginal effect of the interaction term, can be of opposite sign, and its statistical significance is not calculated by standard software."

26. Our normalized creativity measures are comprised between 0 and 1.

Table 2.6 – Probability of typing in disorder

	(1)	(2)
DT	-.1460 (.7260)	
IT	-.9216* (.5566)	
CI		-1.2266* (.7450)
Risk seeking	.0174 (.0551)	.0178 (.0553)
Age	-.4433** (.2170)	-.4273** (.2105)
Women	-.3026 (.2497)	-.2785 (.2564)
Constant	6.9093** (3.2423)	6.6079** (3.1562)
Pseudo R^2	.1165	.1138
Observations	436	436
Clusters	109	109
AUC	0.73	0.73

Notes: **Probit regression. Standard errors** in parentheses. **Significance level:** * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. **AUC** is the area under the roc curve. **Control variables:** academy, session, order of the sub-task, comprehension errors, and big five personality traits.

Indeed, it appears in Table 2.7 (columns 1 and 2) that free riding hurts the performance of pairs in the typing task unless an IT or CI-gifted subject belongs to the pair. However same table shows that competition annihilates the negative incentives of free-riders in pairs and pushes the latter to cooperate, especially so if they are not IT or CI-gifted.

The consequence of such behavior is that more creative individuals have a higher probability of producing more than the partner (Table 2.8) even if there is no difference on ability, as creative and non-creative individuals perform equally when they are alone - and we control for that. Increasing the score for IT in 0.1 point represents 4.0%²⁷ more chances of being the more productive partner ($p = 0.021$), while an increment for CI of 0.1 points increases in 5.6% the chances to perform better than the partner ($p = 0.032$).

A question remains: *Why should creative people (IT especially) be better cooper-*

27. Average marginal effects.

ators? Maybe because they are rather intelligent and understand the benefits from cooperation in teamwork.

Table 2.7 – Typing task: number of correct codes

	(1)	(2)
DT	-4.0593 (4.4643)	
IT	2.8257 (3.4087)	
CI		.2591 (4.2621)
Pair	-3.9579*** (1.3085)	-3.9609*** (1.2739)
Pair × DT	2.7899 (2.9496)	
Pair × IT	3.6756* (2.2080)	
Pair × CI		6.6006** (2.7286)
Competition	-2.0658 (1.5200)	-2.1093 (1.4991)
Competition × DT	.6241 (3.1807)	
Competition × IT	2.7513 (2.2738)	
Competition × CI		3.8462 (3.4468)
Pair × Competition	5.7920*** (1.8696)	5.8758*** (1.8573)
Pair × Competition × DT	-.8118 (3.8110)	
Pair × Competition × IT	-5.4750* (2.9447)	
Pair × Competition × CI		-7.2545* (3.9394)
Risk seeking	.6195** (.2386)	.6173** (.2428)
Age	-1.1025 (1.4557)	-1.2591 (1.4746)
Women	.3450 (1.3249)	.1878 (1.3048)
Constant	38.1335* (21.7886)	40.7420* (21.9240)
R^2	.2473	.2354
Observations	436	436
Clusters	109	109
AUC		

Notes: **Standard errors** in parentheses. **Significance level:** * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. **Control variables:** academy, session, order of the sub-task, disorder, comprehension errors, and big five personality traits.

Table 2.8 – Probability of producing more than the partner

	(1)	(2)
DT	.2495 (.6889)	
IT	1.1436** (.5165)	
CI		1.5944** (.7723)
Competition	.0568 (.2226)	.0548 (.2224)
Knows partner	.1516 (.1804)	.1462 (.1806)
Disorder	-.1311 (.2263)	-.1434 (.2274)
Risk seeking	.1201** (.0473)	.1192** (.0478)
Age	.1984 (.2494)	.1785 (.2490)
Women	.6840*** (.2398)	.6690*** (.2384)
Constant	-3.1307 (3.9010)	-2.7705 (3.8769)
Pseudo R^2	.1161	.1126
Observations	218	218
Clusters	109	109
AUC	0.72	0.72

Notes: **Probit regression. Standard errors** in parentheses. **Significance level:** * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. **AUC** is the area under the roc curve. **Control variables:** academy, session, order of the sub-task, comprehension errors, and big five personality traits.

Table 2.9 – Buttons task: number of points (OLS)

	Equal solution			Best button			Best moving strategy		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
OLS with robust standard errors									
DT	4.0177 (19.7463)	-45.1557 (38.2898)	-0715 (.1584)	190.6424* (97.1772)	276.6442 (189.4699)		-1.9826 (101.3398)	-339.9120* (199.2368)	
IT	-27.3306** (12.7246)	-67.5949** (29.6202)	.4550 (.4149)	62.0290 (70.2642)	132.6012 (163.3056)		-23.0181 (66.3261)	-300.9659* (168.7388)	
DT × IT		110.2810 (74.7229)		-193.6654 (377.7886)				761.0651 (413.9871)	
CI			-33.3231* (19.8864)			211.8484* (110.1295)			-32.0681 (113.7970)
Time (seconds)	-0.802 (.1558)	-0.845 (.1609)	-0715 (.1584)	1.8347** (.7737)	1.8272** (.7871)	1.8985** (.7561)	5.5016** (1.0859)	5.4581*** (1.0757)	5.5238*** (1.0652)
Number of Switches	.4678 (.4094)	.4777 (.4093)	.4550 (.4149)	9.0381*** (1.9243)	9.0571*** (1.9245)	8.8781*** (1.8730)	9.3714*** (2.0589)	9.3012*** (2.0594)	9.3995*** (2.0553)
Number of Switches 2	-0.049 (.0049)	-0.048 (.0048)	-0.045 (.0049)	-1.1111*** (.0192)	-1.1116*** (.0192)	-1.088*** (.0186)	-1.1189*** (.0218)	-1.1168*** (.0219)	-1.1191*** (.0217)
Competition	-17.0894* (9.1578)	-16.6772* (9.1627)	-15.5345* (8.7424)	-6.4625 (42.1836)	-7.3295 (42.3275)	.2403 (41.6002)	-43.6111 (35.5263)	-40.4956 (34.1625)	-42.7740 (35.5267)
Risk seeking	1.5651 (1.7296)	1.8316 (1.7737)	1.5001 (1.7287)	-6.8083 (8.6250)	-7.3108 (8.6448)	-7.0619 (8.6557)	8.1479 (7.0128)	10.0390 (6.9376)	8.1058 (6.9827)
Age	-4.5314 (5.3362)	-5.5163 (5.4099)	-4.3031 (5.3771)	30.4647 (26.5566)	32.1743 (27.4693)	31.8222 (25.5321)	-18.4636 (24.3861)	-25.0841 (23.9830)	-18.2601 (24.3603)
Women	-4.5308 (6.4773)	-4.3845 (6.4729)	-4.7357 (6.3863)	-37.0883 (32.8394)	-37.1956 (33.0875)	-38.2395 (32.8317)	-73.0544** (34.1956)	-71.3622** (33.7174)	-73.3257** (34.0119)
Constant	530.9237*** (83.6767)	561.4814*** (86.3403)	526.5008*** (84.7226)	-221.0460 (426.1418)	-273.8923 (462.1622)	-245.1969 (412.4857)	579.4138 (389.7863)	789.9501** (391.1686)	575.1044 (389.7113)
R^2	.1480	.1574	.1401	.3652	.3663	.3604	.4440	.4574	.4438
Adjusted R^2	.0129	.0169	.0105	.2645	.2606	.2640	.3557	.3670	.3600
Observations	169	169	169	169	169	169	169	169	169

Notes: **Standard errors** in parentheses. **Significance level:** * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. **Control variables:** academy, session, order of the sub-task, disorder, incorrect codes, comprehension errors, and big five personality traits.

3.2. Buttons: a creative exploratory task

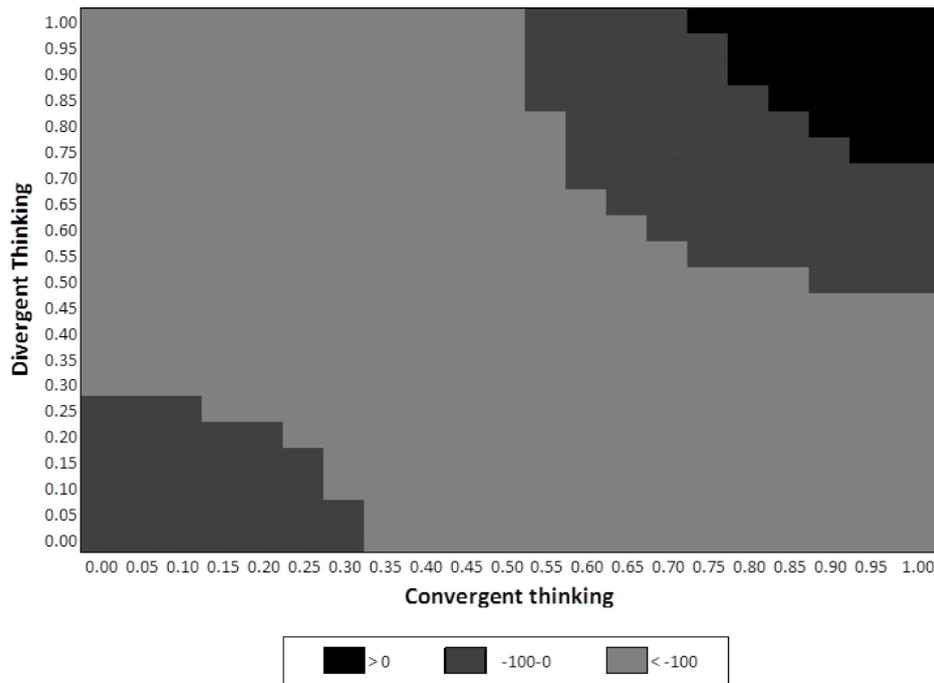
Result 4: Divergent thinkers are more productive than others in research activities. We then turn to the three buttons subtasks in Table 2.9. The first subtask offered no best button at any stage of the game. All buttons yielded equal scores on average and the best strategy ex-ante was to click the same button all the time. However, players were not aware that it was no worth searching and it took them some time searching before they discovered the truth. Uncreative players would then score better -or no worse- than more creative ones. The results displayed in columns (1), (2) and (3) show exactly this: adolescents with high IT and CI scores performed worse on average than their less creative classmates. We argue that integrative thinkers fail trying to "converge" to a solution that does not exist for this subtask.

The most important result appears in columns (4) and (6). Subjects with higher scores for DT and CI performed the "best buttons task" better than less creative ones, and this is mainly caused by their higher ability in DT. Since the best buttons task was designed to measure the efficiency of exploration, such result means that DT and CI correlate with search behavior (assumption H2). Note that the coefficient for DT is still positive when including the interaction term between DT and IT, but it is not significant at 10% level ($p = 0.146$).

By contrast, creativity measures don't correlate *separately* with entrepreneurial behavior since creativity scores don't enhance the performance in the "best moving strategy" - columns (7) and (9). Thus the psychological approach to creativity overlooks the entrepreneurial type of creativity (assumption H3 revised). When analyzing the interaction between DT and CT in Table 2.9 column (8), there is an evidence that a combination of high scores for both DT and IT captures the entrepreneurial type of creativity. Figure 2.2 shows that the positive effect of the interaction between DT and IT in the "best moving strategy series" is really limited (black area), and that a high score in one of the measures and a low score in the other (light gray area) predicts

the opposite, these individuals are not at all classified in the entrepreneurial type of creativity. We do not have a single participant placed in the black area²⁸, and this is the reason why our creativity index (CI) does not capture the entrepreneurial type of creativity.

Figure 2.2 – Interaction between DT and CT: additional points in the "best moving strategy" series.



Looking deeper the determinants of performance in the buttons task we observe that the time spent in the task and the number of button switches have an important impact over results in the "best button" and "best moving strategies" series, but no impact over the "equal solution" series (Table 2.9). This result is consequence of (and validates) the design of the task since the time spent is a proxy of the "thinking effort" in the task. More time in the task can drive to higher chances of finding the "solution" that gives more points, which is valid for "best button" and "best strategy", but has

28. Figure 2.3 shows the distribution of our participants according to their scores on DT and IT.

the inverse impact for "equal solution" (negative coefficient, but not significant). Note that the solution to the "best moving" is more complex and the coefficient for "time" is more important in this sub-task when compared to "best button", it explains why you need both DT and IT to succeed the best moving strategy. The number of switches is also positively correlated to the outcome, but it is not a linear function: once reached the optimal number of switches, the impact over the performance becomes negative²⁹. As expected there is no impact of the number of switches over the "equal solution" series. It is important to highlight that participants do not spend more time in the task because they switch more since the correlation between this two variables is low and not significant, except for "equal solution"³⁰.

We have two important evidences that the "best moving strategy" series measures the entrepreneur type of creativity. First, the vocational trait "Enterprising" predict positively the performance on this sub-task³¹. Second, women had a worse performance on this sub-task and not on two other buttons series (Table 2.9) reflecting the existing literature on gender differences and entrepreneurship: the rate of entrepreneurship for womens is lower than for men (Syed et al., 2010).

Result 5: Research type of creativity increases the chance to capture unexpected rents. One important characteristic of this task is the unexpected rent for those that are in the correct path: when finding the correct button or strategy, participant is rewarded with a jackpot³². However, the correct button remains the

29. Optimum number of switches is 48, 41 and 39 in the "equal solution", "best button" and "best moving strategies" respectively. It corresponds to 11.16, 183,62 and 184,61 points for "equal solution", "best button" and "best moving strategies" respectively. The optimal number of switches is much more important for the "best button" and "best moving strategies", given the potential outcome (additional number of points) and the average score for each buttons series (Table 2.16). We have only one participant with the optimal number of switches for the "best strategy", nobody for other series.

30. Pearson's correlation coefficient between Time and Switch: "equal solution": $r = 0.2975$, $p = 0.000$; "best button": $r = 0.1127$, $p = 0.1509$; "best moving strategy": $r = -0.0200$, $p = 0.7994$.

31. This regression is presented in the Appendix to this Chapter. See Table 2.23 column (3), for more details.

32. It is limited to "best button" and "best moving strategy" series.

best even if the player misses the jackpot. Table 2.10 displays results *without* jackpots' additional points and reproduces the patterns found in Table 2.9, obviously with lower coefficients: scores for DT and CI³³ predict positively the performance on best button series (columns (1) and (3)), and a high scores on *both* DT and IT is needed to perform better in the best moving strategy series (column (5)).

Table 2.11 shows the effect of creativity on the probability of finding the jackpot. This effect is task-dependent, creativity helps "making a fortune" with the best buttons task that requires high searching abilities, but it is has no effect over the best moving task that requires entrepreneurial abilities, apparently of a different nature. Whereas the researcher must combine reflection time and randomness (switching) to discover a hidden jackpot, the entrepreneur must exploit tenaciously a beneficial innovation and shift to a better opportunity only when profits decline. For the best buttons task, chances of finding the jackpot are 3.7% ($p = 0.049$), 2.9% ($p = 0.027$), and 6.3% ($p = 0.001$) higher when increasing by 0.1 point the scored for DT, IT and CI respectively.

Note that the time spent in the task is always positively correlated to the chances of finding the hidden jackpot, showing that outcomes in those tasks are not random. Finding the best solution requires thinking before clicking.

3.3. Satisfaction

Result 6: creative subjects are engaged in costly search when performing creative tasks, but not when performing simple repetitive tasks. Participants are asked about their satisfaction to perform the typing and buttons³⁴ task on a 10-point scale. We asked them to report performance satisfaction and task satisfaction

33. The coefficient for CI is still positive without the jackpot, but not significant at 10% level ($p = 0.280$)

34. Participants report their satisfaction two times, after the buttons task (representing the satisfaction to the 3 buttons series) and after the typing task (representing the satisfaction to the 4 typing series)

Table 2.10 – Buttons task (no jackpot)

OLS with robust standard errors	Best button			Best moving strategy		
	(1)	(2)	(3)	(4)	(5)	(6)
DT	114.2889* (67.4455)	164.3650 (121.6478)		-14.9656 (62.1987)	-227.4805* (120.7669)	
IT	3.2245 (44.0508)	44.3164 (107.9555)		-4.9317 (40.4707)	-179.7258* (102.6128)	
DT × IT		-112.7650 (253.8242)			478.6138* (246.2304)	
CI			82.2606 (75.9178)			-16.5261 (69.6122)
Time (seconds)	1.0819** (.4787)	1.0775** (.4864)	1.1370** (.4627)	3.1053*** (.6466)	3.0779*** (.6409)	3.0947*** (.6314)
Number of Switches	4.2917*** (1.2111)	4.3027*** (1.2110)	4.1535*** (1.1705)	5.3166*** (1.1739)	5.2725*** (1.1785)	5.3032*** (1.1758)
Number of Switches ²	-.0544*** (.0120)	-.0547*** (.0120)	-.0524*** (.0114)	-.0675*** (.0124)	-.0661*** (.0126)	-.0674*** (.0124)
Competition	-1.2816 (27.4811)	-1.7864 (27.5539)	4.5067 (27.5294)	-19.6778 (21.2228)	-17.7185 (20.2748)	-20.0771 (21.2464)
Risk seeking	-2.1663 (5.2156)	-2.4588 (5.1802)	-2.3852 (5.2316)	3.0248 (4.2636)	4.2141 (4.1649)	3.0449 (4.2441)
Age	17.7050 (16.2628)	18.7004 (16.9204)	18.8773 (15.6463)	-20.7847 (14.9175)	-24.9481* (14.7678)	-20.8817 (14.8713)
Women	-31.6413 (20.8070)	-31.7038 (20.9849)	-32.6354 (20.8393)	-36.2268* (21.7020)	-35.1627 (21.3225)	-36.0974* (21.6113)
Constant	-23.6597 (256.3850)	-54.4303 (282.9282)	-44.5153 (248.6750)	495.2801** (234.8235)	627.6808*** (237.9558)	497.3356** (234.6439)
R^2	.3295	.3305	.3201	.4209	.4359	.4208
Adjusted R^2	.2232	.2189	.2176	.3291	.3419	.3336
Observations	169	169	169	169	169	169

Notes: **Standard errors** in parentheses. **Significance level:** * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Control variables: academy, session, order of the sub-task, comprehension errors, and big five personality traits.

separately.

Table 2.12 reports the mean value for each type of satisfaction³⁵. We observe that on average participants are equally satisfied by their performance in both tasks, but they report more satisfaction with the typing task than with the buttons task. We believe that, although retyping codes seems to be less attractive than finding a solution in the buttons task, it is also less stressful to risk-averse subjects and working in pairs and seeing the partner's photo pleased adolescents.

H4 is confirmed in Table 2.13. Creativity scores are not significantly correlated to task satisfaction for typing that was simple and repetitive, but they are negatively

35. Due to technical issues we did not record the values for satisfaction in the buttons task during the first five experimental sessions. Results are not different from those reported in Table 2.12 when considering the same sample in both tasks.

Table 2.11 – Finding the jackpot

	Best button		Best moving strategy	
	(1)	(2)	(3)	(4)
DT	1.4032*		.2168	
	(.7302)		(.5968)	
IT	1.1066**		-.3465	
	(.5186)		(.4039)	
CI		2.3958***		-.3111
		(.7785)		(.6772)
Time (seconds)	.0163**	.0164**	.0349***	.0352***
	(.0066)	(.0066)	(.0078)	(.0077)
Number of Switches	.0815***	.0812***	.0502***	.0507***
	(.0151)	(.0150)	(.0127)	(.0167)
Number of Switches 2	-.0010***	-.0010***	-.0006***	-.0006***
	(.0002)	(.0002)	(.0001)	(.0001)
Competition	-.0465	-.0343	-.3658	-.3437
	(.3148)	(.3085)	(.2400)	(.2390)
Risk seeking	-.0916	-.0920	.0643	.0639
	(.0657)	(.0660)	(.0452)	(.0452)
Age	.2141	.2157	.0165	.0203
	(.2136)	(.2113)	(.1415)	(.1420)
Women	-.1844	-.1899	-.5150***	-.5150***
	(.2522)	(.2519)	(.1883)	(.1898)
Constant	-5.2559	-5.2730		
	(3.3779)	(3.3536)		
Cut 1			-.6049	-.5000
			(2.3210)	(2.3332)
Cut 2			.5671	.6704
			(2.3270)	(2.3378)
Cut 3			1.5159	1.6173
			(2.3324)	(2.3421)
Pseudo R^2	.3120	.3116	0.1853	0.1841
Observations	169	169	169	169

Notes: **Probit** regression for Best button. **Ordered probit** regression for Best moving strategy since there are three possible jackpots in this subtask. **Standard errors** in parentheses. **Significance level:** * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. **Control variables:** academy, session, order of the sub-task, comprehension errors, and big five personality traits.

Table 2.12 – Satisfaction

Satisfaction	Typing	Buttons	Sign.
with task	7.48	6.77	*
with performance	6.27	6.52	ns
Sign.	***	ns	
n	109	71	

Notes: **Wilcoxon significance test:** * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, *ns* non significant at 10% level.

and significantly correlated for the buttons task and for all scores of creativity³⁶. The *investment* in search by creative subjects comes at a cost which is measured by their

36. For more information, see the Table 2.24 Appendix to this Chapter. It reproduces the results here presented considering the same sample for both tasks. Results are globally similar.

significant short-run loss of satisfaction in columns (3) and (4) even though this cost is (partially) mitigated by the pleasure of a high performance.

Table 2.13 – Task satisfaction

OLS with robust standard errors	Typing		Buttons	
	(1)	(2)	(3)	(4)
DT	5.9847 (3.8682)		-15.4860** (6.7138)	
IT	-1.6349 (2.5692)		-4.6131 (3.1464)	
CI		1.7954 (3.4610)		-15.1664*** (4.9155)
Performance satisfaction	.5673** (.2176)	.5040** (.2145)	-.5709* (.3205)	-.4653 (.3067)
Performance satisfaction × DT	-.6576 (.5173)		2.3700** (.9097)	
Performance satisfaction × IT	.1355 (.3778)		.4135 (.5049)	
Performance satisfaction × CI		-.2362 (.4901)		2.1596*** (.6522)
Risk seeking	-.1543 (.1058)	-.1481 (.1057)	-.5867*** (.1341)	-.5391*** (.1324)
Age	.7306* (.4300)	.7841* (.4288)	1.1110** (.4542)	1.0528** (.4571)
Women	.0883 (.4915)	.1973 (.4880)	-1.2580** (.5758)	-1.4052** (.5729)
Constant	-7.9911 (6.6904)	-8.5234 (6.6852)	-2.8585 (7.8296)	-3.4258 (7.8570)
R^2	.2935	.2737	.6338	.6130
Adjusted R^2	.1427	.1380	.5163	.5074
Observations	109	109	71	71

Notes: **Standard errors** in parentheses. **Significance level:** * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. **Control variables:** academy, session, order of the sub-task, comprehension errors, and big five personality traits.

4. Creativity and school achievements

We just identified how real-effort tasks in a laboratory are influenced by creativity. Schooling achievement is of another kind: it depicts "real life" performance as measured by the educational system and it is a major determinant of future market productivity and earnings. We asked the 10th graders to report their *Brevet* grades: general average, Mathematics, French, History/Geography and Art history grades. Unfortunately, around one third of them did not get their grade book at the time of the experiment and could not respond. The French grading system is based on a 20 points scale. Half

of the grade in Math, French and History/Geography comes from year-long evaluation, the other half comes from the final exam. The Art history grade relies on a report and an oral presentation. Several studies established a high correlation between reported and actual grades³⁷.

We regress below grades in Math, French, and the general grade³⁸ obtained at the *Brevet* on individual measures of creativity, personality, general risk tolerance, age and gender, using OLS. As the two components of creativity, that is, integrative and divergent thinking, may have different and nonlinear effects on school achievement, we introduced them separately and with an interaction term. Then, we added the individual scores in the “best button task” and in the “best moving strategy task” as different and partly independent measures of economic creativity. After trials and errors, we also decided to dis-aggregate three of the five personality variables (agreeableness, emotional stability, and openness) into their positive and negative components because these are not strictly opposite within the condensed TIPI score and don’t necessarily exert opposite effects. We present our preferred regressions in Table 2.14, controlling for the academic districts.

The score for the best moving strategy task was never significant and was removed from Table 2.14. It is not too surprising that the middle school grades don’t screen the entrepreneurial ability at this early stage of life. The impact of creativity on the three grades reported here is contrasted. For instance, integrative and divergent thinking appear to be strong complements³⁹ in Math but substitutes elsewhere (negative interaction term). Divergent thinking, associated with imagination, is beneficial in French literature and in the general grade but it is harmful in Math in the absence of con-

37. Students who have lower grades tend to misreport (over-estimate) more their self-reported grades (Cole, Rocconi and Gonyea, 2012).

38. We don’t present here regressions for History/Geography and Art history because there are no effects of creativity over grades in these subjects. Estimations for History/Geography and Art history are presented in the Appendix to this Chapter (Table 2.25).

39. Even if DT is not significant at 10% level ($p = 0.376$) in column (1), the coefficient for CI (column (3)) - that is the average between DT and IT - is significant ($p = 0.022$) and higher than the sum of DT and IT coefficients in column (1).

vergent thinking. Random search behavior and risk tolerance is appreciated in French where students are expected to have a broad look and to develop controversial ideas; but it has no impact in Math and for the general grade. Interestingly, conscientiousness has no significant effect on each specific grade but the attention uniformly devoted to all fields makes a difference on the general grade. Anxiety is most harmful to the mathematical ability, probably because it inhibits the quest for rigor. And surprisingly, for 9th graders at least, a conventional uncreative attitude is rewarded in French while a curious but complex personality is not. Lastly, girls have lower grades than boys in Math and higher grades in French even after controlling for a rich set of creativity and personality measures.

Table 2.14 – OLS regression of school achievements of French 9th graders

	Maths			French			General grade		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
OLS with robust standard errors									
DT	1.6207 (1.7241)	.3647 (4.0842)		.2644 (1.3467)	7.0050** (3.1795)		.5644 (1.0635)	6.3522** (2.4708)	
CT	2.7804** (1.2541)	1.7649 (3.2453)		1.4983 (.9912)	6.9670*** (2.5400)		.4612 (.7891)	5.1939** (1.9889)	
DT × CT		2.6083 (7.6817)			-13.8787** (5.9569)			-11.9462** (4.6308)	
CI		4.7671** (2.0457)				2.1338 (1.6023)			.9961 (1.2524)
Best button (points)	-0.001 (.0014)	-0.001 (.0014)	-0.002 (.0014)	.0027** (.0011)	.0025** (.0011)	.0026** (.0011)	.0008 (.0009)	.0006 (.0009)	.0008 (.0009)
Academy disctrict (ref. Paris)									
Créteil	1.5082** (.7490)	1.4898* (.7542)	1.5841** (.7337)	1.1120* (.5843)	1.2422** (.5748)	1.1838** (.5750)	1.0327** (.4524)	1.0845** (.4408)	1.0260** (.4419)
Versailles	1.5939** (.7172)	1.5525** (.7305)	1.6248** (.7126)	1.0360* (.5721)	1.2772** (.5697)	1.0652* (.5695)	1.0786** (.4560)	1.2733** (.4502)	1.0782** (.4538)
Risk seeking	-0.0763 (.1626)	-0.0701 (.1643)	-0.0724 (.1619)	.2671** (.1277)	.2331* (.1259)	.2733** (.1271)	-0.0092 (.1005)	-0.0233 (.0980)	-0.0101 (.0994)
Age	-0.1517 (.5368)	-0.1735 (.5429)	-0.1449 (.5348)	-0.0670 (.4126)	-0.0899 (.4096)	-0.0704 (.4117)	-0.1074 (.3386)	-0.1066 (.3328)	-0.1066 (.3368)
Women	-1.1109* (.6005)	-1.0966* (.6045)	-1.1359* (.5967)	1.6375*** (.4684)	1.5439*** (.4604)	1.6197*** (.4668)	1.902 (.3685)	1.002 (.3604)	1.930 (.3650)
Personality traits									
Conscientiousness	.2558 (.2828)	.2440 (.2862)	.2586 (.2818)	.0029 (.2259)	.0584 (.2224)	.0085 (.2253)	.3800** (.1703)	.4301** (.1669)	.3804** (.1694)
Extraversion	.0258 (.2889)	.0304 (.2905)	.0380 (.2871)	.0648 (.2210)	.0418 (.2166)	.0709 (.2203)	.0589 (.1694)	.0424 (.1650)	.0577 (.1679)
Critical, quarrelsome	.4740** (.2285)	.4583* (.2341)	.4931** (.2251)	-0.2020 (.1794)	-0.1171 (.1794)	-0.1800 (.1766)	.1826 (.1421)	.2242 (.1392)	.1816 (.1407)
Sympathetic, warm	-0.0677 (.2421)	-0.0592 (.2444)	-0.0698 (.2412)	-0.2879 (.1908)	-0.3391* (.1881)	-0.2888 (.1903)	-0.2324 (.1467)	-0.2733* (.1437)	-0.2326 (.1460)
Anxious, easily upset	.0680 (.2025)	.0512 (.2093)	.0844 (.1997)	.0478 (.1594)	.1393 (.1609)	.0641 (.1575)	.0088 (.1164)	.0639 (.1153)	.0076 (.1147)
Calm, emotionally stable	-0.3663** (.1597)	-0.3534** (.1648)	-0.3584** (.1585)	-0.0855 (.1269)	-0.1481 (.1271)	-0.0814 (.1265)	-0.0689 (.0963)	-0.1355 (.0972)	-0.0694 (.0956)
Open to new experiences, complex	-0.0965 (.1969)	-1.002 (.1981)	-0.9958 (.1962)	1.498 (.1559)	1.1654 (.1528)	1.492 (.1555)	.2360** (.1181)	.2689** (.1156)	.2357** (.1174)
Conventional, uncreative	-0.0428 (.1896)	-0.0530 (.1928)	-0.0283 (.1872)	.3153** (.1446)	.3726** (.1437)	.3259** (.1436)	.1851 (.1182)	.2482** (.1176)	.1841 (.1169)
Constant	10.3023 (8.7691)	11.1878 (9.1850)	9.8086 (8.6939)	8.4940 (6.7627)	3.1581 (7.0067)	8.1750 (6.7344)	10.0736* (5.4824)	5.6853 (5.6010)	10.0949* (5.4489)
R ²	.2450	.2459	.2428	.3087	.3436	.3050	.2176	.2660	.2176
Adjusted R ²	.1277	.1202	.1336	.2013	.2342	.2047	.0949	.1424	.1036
Observations	119	119	119	119	119	119	119	119	119

Notes: **Standard errors** in parentheses. **Significance level:** * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. **Personality traits:** Agreeableness: Critical, quarrelsome (reverse-scored item) + Sympathetic, warm; Emotional Stability: Anxious, easily upset (reverse-scored item) + Calm, emotionally stable; Openness to Experiences: Open to new experiences, complex + Conventional, uncreative (reverse-scored item)

5. Conclusion

From an economic perspective, creative behavior must be judged by individual's propensity to innovate in production (and consumption) activities, distinguishing two types of economic innovators: *researchers* and *entrepreneurs*. We propose an economic experiment with two real-effort tasks to observe the performance of creative individuals in production, using three psychological measures of creativity: the graphical divergent thinking, the graphical integrative thinking and an aggregated creativity index.

We find that divergent *thinkers* are more productive than others in exploration activities, such result means that the psychological assessment of divergent thinking correlates with the *researcher* type of economic innovator. Otherwise, the *entrepreneurial* type was not identified among our scores of creativity. We found an evidence that entrepreneurs are those with high scores for both divergent thinking and convergent thinking, but we do not have any participant with these characteristics in our sample given the low correlation between these measures ($r = 0.13$). This result suggests that entrepreneurship, the ability to capture unexpected rents, may be related to other individual characteristics such as risk preferences and tenacity but not to creativity as we expected.

Additionally, we observe that creative individuals are no more productive than others in repetitive tasks, but they behave differently than less creative individuals in this type of task: integrative *thinkers* are more cooperative when working in pairs, maybe because they are intelligent and understand the benefit of cooperation in teamwork. Indeed, the idea that creative individuals are intelligent is reinforced by the performance at school. Creativity scores play an important role on school achievements, they are positively correlated to grades on Maths, French and the general grade for 10th graders.

Further research must address in what extent these evidences are related to creativity or to subprocesses involved in the two aspects of creative production described

before - such as flexibility, divergent thinking, selective encoding, associative thinking and selective comparison.

The present exploration into this new field is preliminary but it suggests that creative potential would be a valuable addition to common personality traits and IQ for predicting school achievement and future earnings. However, progress in the use of creativity in economic research requires simple and robust measures which are still to come. Besides, an aggregated creativity index is not always warranted as it may mask on some occasions the offsetting, or very unequal, effects of its main components, that is, divergent thinking and integrative or convergent thinking.

Bringing the economic perspective to the analysis and measurement of creativity has shown that the entrepreneurial type of creativity so prominent in the economic theory of innovation and growth is currently overlooked in psychological measures. Further research is needed to develop simple, easily replicable tasks like our buttons tasks, to identify pure types of creative behavior and make use of the resulting creativity scores to test the impact of creativity in various contexts.

A. Appendices

A.1. Tables and figures

Table 2.15 – Descriptive statistics

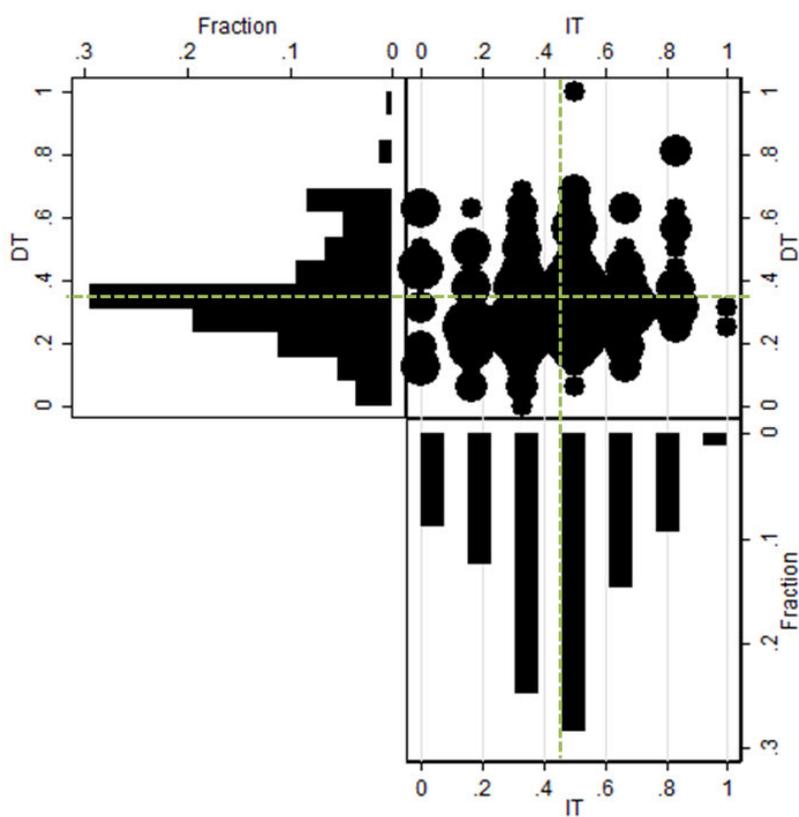
Variable	Mean	Std. Dev.	n
Demographics			
Women	0.46	–	169
Age	15.18	0.57	169
Personality (scale 1-7)			
Extraversion	4.21	1.33	169
Openness	4.99	1.19	169
Agreeableness	5.41	1.05	169
Conscientiousness	5.15	1.15	169
Emotional stability	4.56	1.38	169
Vocation (scale 0-4)			
Realistic (doer)	1.31	0.80	169
Investigative (thinker)	1.33	1.02	169
Artistic (creator)	1.80	1.06	169
Social (helper)	1.37	0.92	169
Conventional (organizer)	1.42	0.96	169
Enterprising (persuader)	1.64	0.90	169
Risk taking (scale 1-10)			
Risk general	6.60	2.10	169
Risk health	4.24	3.29	169
Risk leisure	7.03	2.58	169
Risk others	4.10	3.10	169
Risk school	5.21	2.86	169

Table 2.16 – Variables description: buttons task ($n = 169$)

Variable	Description	Equal solution		Best button		Best strategy	
		Mean	SD	Mean	SD	Mean	SD
Points	Total score: sum of 100 clicks' points.	393.73	40.55	360.52	214.85	460.21	232.75
Points "no-jackpot"	Total points not considering jackpots' points.	393.73	40.55	274.13	132.07	290.39	138.93
Jackpot	Number of jackpots found on average. Jackpots are limited to best button (1 jackpot of 100points) and best strategy (3jackpots of 100 points each) series.	—	—	.43	.50	1.70	1.06
Switches	Number of buttons switches.	39.04	34.36	40.94	34.75	32.16	28.35
Time	Time spent in each buttons series	49.26	28.37	41.42	17.87	42.80	16.38

Note: **Points**: the maximum and minimum number of possible points, if participants select the button with the highest/lowest outcome for each click, is (max, min): Equal solution = (1110, -277), Best button = (912, -82), and Best strategy = (926, -63).

Figure 2.3 – Distribution of scores: DT and IT.



Notes: the size of the spots increase with the observations they contain. Mean scores for DT and IT are represented by the green slashed lines.

Table 2.17 – Variables description: typing task ($n = 109$)

Variable	Description	All		A-N		A-C		P-N		P-C	
		Mean	SD								
Correct codes	Number of codes correctly typed.	22.28	6.99	22.31	7.29	21.59	6.23	20.63	6.94	24.58	6.93
Incorrect codes	Number of codes incorrectly typed.	1.47	1.07	1.43	.93	1.60	1.35	1.47	1.07	1.40	.87
Efficiency	Ratio between correct codes and total typed.	.93	.06	.93	.05	.93	.06	.92	.08	.94	.04
n_disorder	Number of codes picked randomly, i.e. no the sequence of code typed before.	5.07	9.09	4.66	8.52	5.01	8.73	5.25	9.43	5.36	9.73
Disorder	Dummy variable: 1 if typed at least one code in disorder; 0 if typed all codes in order. Measure for each typing sub-task.	.30	.46	.31	.46	.31	.46	.30	.46	.29	.46
Knows partner	Dummy variable: 1 if knows the partner; 0 if doesn't. Limited to pair condition.	.38	.49	—	—	—	—	.37	.48	.40	.49
More than partner	Dummy variable: 1 if performance is higher than partner; or if performance is equal or inferior.	.47	.50	—	—	—	—	.45	.50	.49	.50

Notes: All: all sample; A-N: alone/no-competition; A-C: alone/competition; P-N: pair/no-competition; P-C: pair/competition.

Table 2.18 – Other variables description

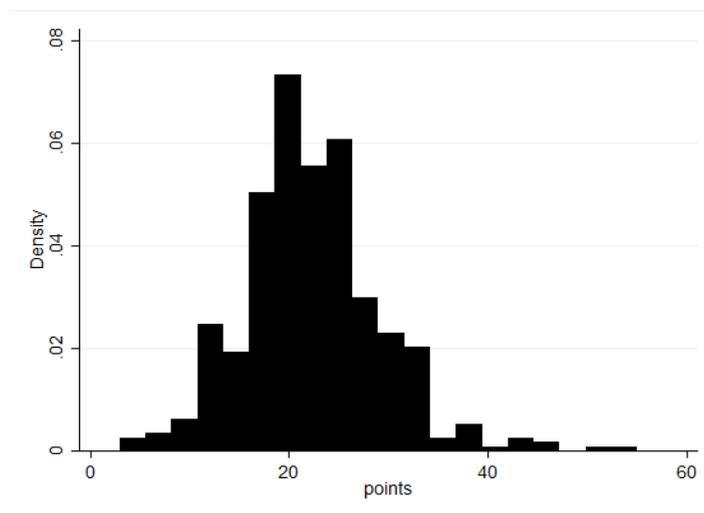
Variable	Task	Description	Mean	n
Comprehension errors	Typing	Dummy variable: 1 if make errors in comprehension questions; 0 if don't.	.50	109
Comprehension errors	Buttons	Dummy variable: 1 if make errors in comprehension questions; 0 if don't.	.43	169
Competition	Buttons	Part of participants under competition	.51	109

Table 2.19 – Number of correct codes in the typing task

	Mean	Median	Std. Dev.	Skewness	Kurtosis	n
Correct codes	22.28	22	6.99	0.59	4.67	436

Notes: number of correct codes consolidated for the 4 typing series, for the total of 109 participants.

Figure 2.4 – Distribution of correct codes in the typing task



Notes: number of correct codes consolidated for the 4 typing series, for the total of 109 participants.

Table 2.20 – Grades: descriptive statistics ($n = 119$)

Subject	Mean	SD
General grade	12.70	1.93
Math	10.51	3.18
French	12.01	2.63
History/Geography	13.32	2.41
Art history	14.94	3.33

Table 2.21 – Correlations between creativity measures and personality variables.

Characteristics	DT		IT		CI	
	<i>r</i>	<i>p</i> – value	<i>r</i>	<i>p</i> – value	<i>r</i>	<i>p</i> – value
Women	-0.0434	0.5749	-0.0629	0.4167	-0.0722	0.3512
Age	-0.0162	0.8339	-0.1155	0.1348	-0.0982	0.2039
Personality (scale 1-7)						
Extraversion	0.0873	0.2593	0.1613	0.0362	0.1720	0.0253
Openness	0.1496	0.0523	-0.0301	0.6977	0.0575	0.4578
Agreeableness	0.0401	0.6051	-0.0272	0.7255	0.0006	0.9941
Conscientiousness	0.0036	0.9630	-0.1073	0.1648	-0.0812	0.2942
Emotional stability	-0.0986	0.2021	-0.0729	0.3464	-0.1097	0.1556
Vocation (scale 0-4)						
Realistic (doer)	0.0387	0.6174	0.0449	0.5626	0.0556	0.4725
Investigative (thinker)	0.0069	0.9294	0.1342	0.0819	0.1002	0.1948
Artistic (creator)	-0.0639	0.4092	0.0025	0.9741	-0.0326	0.6742
Social (helper)	-0.1042	0.1776	-0.0328	0.6725	-0.0817	0.2913
Conventional (organizer)	-0.0949	0.2195	0.0056	0.9428	-0.0470	0.5440
Enterprising (persuader)	-0.1227	0.1120	0.0219	0.7772	-0.0493	0.5243
Risk taking (scale 1-10)						
Risk general	0.0896	0.2468	0.1322	0.0866	0.1508	0.0504
Risk health	0.1561	0.0426	0.1282	0.0968	0.1836	0.0169
Risk leisure	0.1509	0.0502	0.1258	0.1033	0.1789	0.0200
Risk others	0.0073	0.9253	-0.0368	0.6346	-0.0246	0.7510
Risk school	0.0835	0.2806	0.0243	0.7537	0.0639	0.4090

Note: *r*: Pearson's correlation coefficient.

Table 2.22 – Typing task: outcomes and efficiency

Regression:	Ordinary least squares		Negative binomial		Poisson	
DT	-1.9409 (4.0758)		-.0906 (.1744)		-.0845 (.1763)	
IT	4.4303 (3.3644)		.1918 (.1449)		.1990 (.1484)	
CI		3.8464 (4.5047)		.1653 (.2019)		.1757 (.2003)
Pair	-1.2061** (.5128)	-1.2022** (.5134)	-.0553** (.0224)	-.0549** (.0224)	-.0536** (.0224)	-.0536** (.0224)
Competition	-.3815 (.5437)	-.3749 (.5414)	-.0138 (.0237)	-.0131 (.0236)	-.0154 (.0236)	-.0152 (.0236)
Pair × Competition	2.8253*** (.6810)	2.8142*** (.6817)	.1212*** (.0300)	.1200*** (.0299)	.1207*** (.0300)	.1202*** (.0300)
Incorrect codes	-.9005* (.5077)	-.9255* (.5171)	-.0425* (.0238)	-.0437* (.0243)	-.0426* (.0250)	-.0441* (.0255)
Disorder		-.4753 (1.2995)		-.0191 (.0560)		-.0191 (.0581)
Risk seeking	.6371** (.2489)	.6331** (.2516)	.0306*** (.0113)	.0302*** (.0114)	.0302*** (.0114)	.0299*** (.0115)
Age	-1.1371 (1.5114)	-1.2850 (1.5264)	-.0540 (.0636)	-.0602 (.0649)	-.0526 (.0660)	-.0590 (.0672)
Women	.1818 (1.2937)	.0403 (1.2666)	.0048 (.0572)	-.0013 (.0559)	.0077 (.0573)	.0003 (.0559)
Constant	35.8582 (21.8610)	38.4391* (21.9692)	3.7315*** (.9276)	3.8437*** (.9421)	3.7060*** (.9521)	3.8208*** (.9633)
R^2	.2216	.2126				
Adjusted R^2	.1762	.1687				
Pseudo R^2			0.0386	0.0369		
Log pseudolikelihood			-1404.152	-1406.6243	-1439.6593	-1443.9053
Observations	436	436	436	436	436	436
Clusters	109	109	109	109	109	109

Notes: **Standard errors** in parentheses. **Significance level:** * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Disorder: typing in disorder is a dummy variable for each of the four rounds participants performed the task: 0 if participant typed all codes in order and 1 if typed at least one code in disorder. Descriptive statistics for efficiency is presented in Table 2.15. **Control variables:** academy, session, order of the sub-task, comprehension errors, and big five personality traits. **Sample size:** due to technical reasons, we have only 109 adolescents participating in the typing task.

Table 2.23 – Buttons task: vocational scores

	Equal solution (1)	Best button (2)	Best moving strategy (3)
Realistic	-3.1033 (6.2189)	-17.7708 (25.5189)	12.9049 (29.7844)
Investigative	1.5321 (5.1405)	3.7722 (18.5010)	6.1661 (17.7896)
Artistic	-1.4734 (3.7176)	18.9707 (15.6888)	-2.3710 (17.2170)
Social	1.8364 (5.7356)	-38.5027 (24.1663)	-38.6434* (23.2129)
Enterprising	-1.7935 (4.6343)	-17.4514 (26.3668)	39.9609* (22.6154)
Conventional	-.2911 (4.2690)	26.1354 (20.6807)	-11.3036 (19.4850)
Time (seconds)	-.0538 (.1656)	2.1938*** (.7918)	5.5671*** (1.1154)
Number of Switches	.3572 (.4281)	8.2452*** (1.8586)	9.4862*** (2.0237)
Number of Switches ²	-.0035 (.0049)	-.1010*** (.0187)	-.1217*** (.0217)
Competition	-13.7534 (9.7539)	-6.4521 (43.7992)	-39.9300 (35.5351)
Risk seeking	1.6464 (1.8421)	-1.9525 (8.3123)	6.7897 (7.8972)
Age	-2.5730 (5.3451)	28.7748 (27.4542)	-24.3606 (27.3824)
Women	-5.9917 (7.2249)	-32.8900 (38.0083)	-44.8990 (41.3618)
Constant	495.6576*** (82.2916)	-145.1999 (443.1668)	612.1651 (419.7179)
R^2	.1331	.3664	.4588
Observations	169	169	169

Notes: **Standard errors** in parentheses. **Significance level:** * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. **Control variables:** academy, session, order of the sub-task, comprehension errors, and big five personality traits.

Table 2.24 – Satisfaction (same sample of buttons task).

	Typing	
	(1)	(2)
DT	3.0755 (4.5271)	
IT	-4.1766 (2.8580)	
CI		-4.0194 (4.1586)
Performance's satisfaction	.1913 (.2684)	.1035 (.2589)
DT × Performance's satisfaction	-.1557 (.6222)	
IT × Performance's satisfaction	.5398 (.4279)	
CI × Performance's satisfaction		.6780 (.5711)
Risk seeking	-.2011 (.1452)	-.1752 (.1428)
Age	1.0512** (.4807)	1.0560** (.4759)
Women	.2970 (.5905)	.4174 (.5834)
Constant	-10.2443 (7.4198)	-9.6946 (7.3074)
R^2	.3901	.3663
Observations	71	71

Notes: **Standard errors** in parentheses. **Significance level:** * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. **Control variables:** academy, session, comprehension errors, and big five personality traits.

Table 2.25 – OLS regression of school achievements of French 9th graders

OLS with robust standard errors	History/Geography			Art history		
	(1)	(2)	(3)	(4)	(5)	(6)
DT	1.2861 (1.3423)	2.5129 (3.3144)		-1.4075 (1.8679)	3.4079 (4.4232)	
CT	.1066 (1.0019)	1.0658 (2.5725)		-2.509 (1.3716)	3.6157 (3.4997)	
DT × CT		-2.4905 (6.1473)			-9.9866 (8.3188)	
CI			1.0561 (1.6045)			-1.3115 (2.2053)
Best button (points)	.0003 (.0011)	.0003 (.0011)	.0004 (.0011)	.0013 (.0015)	.0010 (.0015)	.0012 (.0015)
Academy district (ref. Paris)						
Créteil	.8606 (.5899)	.8791 (.5940)	.7981 (.5818)	-.3643 (.8088)	-.2913 (.8094)	-.2916 (.7928)
Versailles	1.4790*** (.5586)	1.5156*** (.5680)	1.4455** (.5552)	1.3243* (.7933)	1.4788* (.8021)	1.3518* (.7887)
Risk seeking	.0649 (.1280)	.0597 (.1292)	.0616 (.1276)	.2668 (.1771)	.2463 (.1775)	.2717 (.1762)
Age	-.0432 (.4125)	-.0217 (.4175)	-.0446 (.4115)	.5408 (.5745)	.6387 (.5791)	.5410 (.5725)
Women	-.1497 (.4641)	-.1679 (.4681)	-.1274 (.4619)	.1800 (.6559)	.1133 (.6570)	.1637 (.6529)
Personality traits						
Conscientiousness	.3564 (.2241)	.3648 (.2260)	.3598 (.2236)	.4185 (.3149)	.4635 (.3165)	.4164 (.3138)
Extraversion	.1368 (.2260)	.1336 (.2271)	.1262 (.2250)	-.0013 (.3124)	-.0100 (.3118)	.0079 (.3108)
Critical, quarrelsome	.3187* (.1802)	.3317* (.1838)	.3011* (.1781)	.1377 (.2464)	.1988 (.2511)	.1547 (.2432)
Sympathetic, warm	-.4569** (.1931)	-.4630** (.1945)	-.4572** (.1927)	-.1769 (.2659)	-.2102 (.2668)	-.1775 (.2650)
Anxious, easily upset	.1981 (.1637)	.2111 (.1675)	.1853 (.1623)	-.0928 (.2169)	-.0257 (.2236)	-.0762 (.2137)
Calm, emotionally stable	.1379 (.1246)	.1255 (.1288)	.1341 (.1242)	-.1550 (.1724)	-.2013 (.1763)	-.1492 (.1714)
Open to new experiences, complex	-.0780 (.1569)	-.0765 (.1575)	-.0776 (.1565)	.1908 (.2165)	.2071 (.2165)	.1908 (.2157)
Conventional, uncreative	-.0058 (.1496)	.0037 (.1520)	-.0148 (.1486)	.0503 (.2040)	.0956 (.2070)	.0615 (.2020)
Constant	10.5069 (6.8001)	9.6789 (7.1267)	10.8114 (6.7703)	3.1148 (9.3684)	-.5670 (9.8395)	2.7732 (9.3112)
R^2	.1720	.1733	.1681	.1391	.1505	.1371
Adjusted R^2	.0458	.0382	.0503	.0115	.0156	.0183
Observations	122	122	122	125	125	125

Notes: **Standard errors** in parentheses. **Significance level:** * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. **Personality traits:** Agreeableness: Critical, quarrelsome (reverse-scored item) + Sympathetic, warm; Emotional Stability: Anxious, easily upset (reverse-scored item) + Calm, emotionally stable; Openness to Experiences: Open to new experiences, complex + Conventional, uncreative (reverse-scored item).

A.2. Holland's typology

Table 2.26 – Holland's typology

Holland type	Description	Example of professions
Realistic (Doer)	It involves the manipulation of objects, tools, machines, animals. It promotes realistic activities that strengthen technical skills. It corresponds to people who are predominantly realistic and privileges technical and practical activities.	Farmer, engineer, aircraft controller, electrician...
Investigative (Thinker)	It allows the observation and the systematic and abstract investigations of physical, biological or cultural phenomena. It promotes intellectual activity and leads individuals to develop skills of the same nature. This environment is aimed at people with a predominantly intellectual background. It promotes reflection and observation.	Scientific researcher, biologist, chemist, physicist...
Artistic (Creator)	It involves free, undefined activities, with no rigid framework, with the implementation of creative skills. It leads individuals to engage in artistic activities and to develop skills of the same nature. This environment is aimed at people with an artistic predominance, it is not very structured and leaves an important place to the creativity and the improvisation.	Architect, musician, composer, writer, interior decorator...
Social (Helper)	It involves action on others in order to inform, educate, nurture and help them. It leads individuals to engage in social activities and develop skills of the same type. It is aimed at people with a social preference. It allows to be in relation with different people by promoting the social work.	Psychologist, caseworker, general practitioner, teacher...
Enterprising (Persuader)	It involves action on others, in order to achieve personal or organizational goals. It leads individuals to engage in "management" activities where they must lead others. It leads them to develop managerial skills. This type of environment is aimed at people who are predominantly enterprising. It puts people in a competitive situation and allows them to evolve and take on increasing responsibilities.	Manager, salesperson, business executive, buyer...
Conventional (Organizer)	It involves the manipulation of data that can be very diverse in nature. It leads individuals to engage in conventional activities that develop skills of the same type. This environment is aimed at people who are predominantly conventional and favors work organized by clear rules.	Banker, financial analyst, tax expert, court reporter, book-keeper, clerk...

A.3. Buttons task

Table 2.5 presents the accumulated number of points for each 10 clicks in a given buttons series. "Outcomes" present the possible outcomes for the hidden values for each button, and distribution presents the number of times each outcome can appear if selecting the button.

Figure 2.5 – Accumulated distribution of points among buttons.

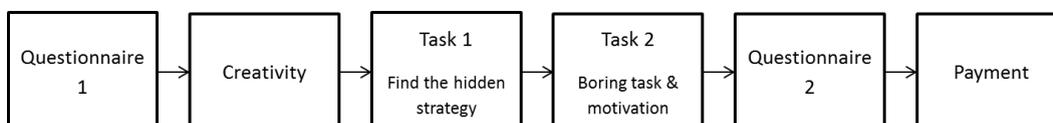
Equal solution Click	Button			
	A	B	C	D
1-10	40	40	40	40
11-20	80	80	80	80
21-30	120	120	120	120
31-40	160	160	160	160
41-50	200	200	200	200
51-60	240	240	240	240
61-70	280	280	280	280
71-80	320	320	320	320
81-90	360	360	360	360
91-100	400	400	400	400
outcomes	(3, 4, 5)	(0, 2, 8)	(-4, 0, 6, 10)	(-8, 0, 16)
distribution (n)	(38, 24, 38)	(23, 36, 41)	(26, 14, 24, 36)	(32, 27, 41)

Best button Click	Button							
	E	F	G	H	I	J	K	L
1-10	8	10	10	12	16	15	10	10
11-20	12	20	62	32	56	45	4	40
21-30	11	30	126	60	64	83	-2	80
31-40	6	40	190	80	80	99	-8	120
41-50	3	50	260	100	112	121	-2	140
51-60	3	60	318	120	136	155	-4	160
61-70	0	70	382	140	152	168	2	180
71-80	1	80	452	160	160	196	-6	220
81-90	2	90	522	180	176	230	8	240
91-100	0	100	760	200	200	250	0	250
outcomes	(-1, 0, 1)	(1)	(0, 1, 7)	(1, 2, 4)	(0, 1, 8)	(0, 1, 3, 7, 9)	(-2, 0, 1, 2)	(0, 1, 10)
distribution	(32, 36, 32)	(100)	(1, 21, 77, 1)	(8, 88, 4)	(68, 8, 24)	(38, 23, 14, 20, 5)	(35, 26, 8, 31)	(66, 10, 24)
jackpot								
value	no	no	200	no	no	no	no	no
"click"	no	no	96	no	no	no	no	no

Best strategy Click	Button			
	M	N	O	P
1-10	2	17	5	6
11-20	2	97	6	19
21-30	2	177	12	29
31-40	10	299	14	41
41-50	90	302	12	52
51-60	170	300	6	64
61-70	302	300	1	73
71-80	302	297	72	83
81-90	302	302	152	93
91-100	300	300	300	100
outcomes	(-2, 0, 1, 8, 100)	(-1, 0, 1, 8, 100)	(-2, -1, 0, 1, 2, 8, 100)	(0, 1, 2)
distribution	(1, 71, 2, 25, 1)	(26, 22, 26, 25, 1)	(13, 14, 17, 20, 10)	(28, 44, 28, ,)
jackpot				
value	100	100	100	no
"click"	65	34	97	no

Experimental Protocol

Figure 2.6 – Order of activities



A.4. Big Five (TIPI)

Here a number of personality traits that may or may not apply to you. Please write a number next to each statement to indicate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other.

Scale: 1: Strongly disagree; 2: Disagree moderately; 3: Disagree a little; 4: Neither agree nor disagree; 5: Agree a little; 6: agree moderately; 7: Strongly agree

I see myself as:

1. Extraverted, enthusiastic
2. Critical, quarrelsome
3. Dependable, self-disciplined
4. Anxious, easily upset
5. Open to new experiences, complex
6. Reserved, quiet
7. Sympathetic, warm
8. Disorganized, careless
9. Calm, emotionally stable
10. Conventional, uncreative

A.5. Creativity scores

Figure 2.7 – Examples of divergent thinking graphical test

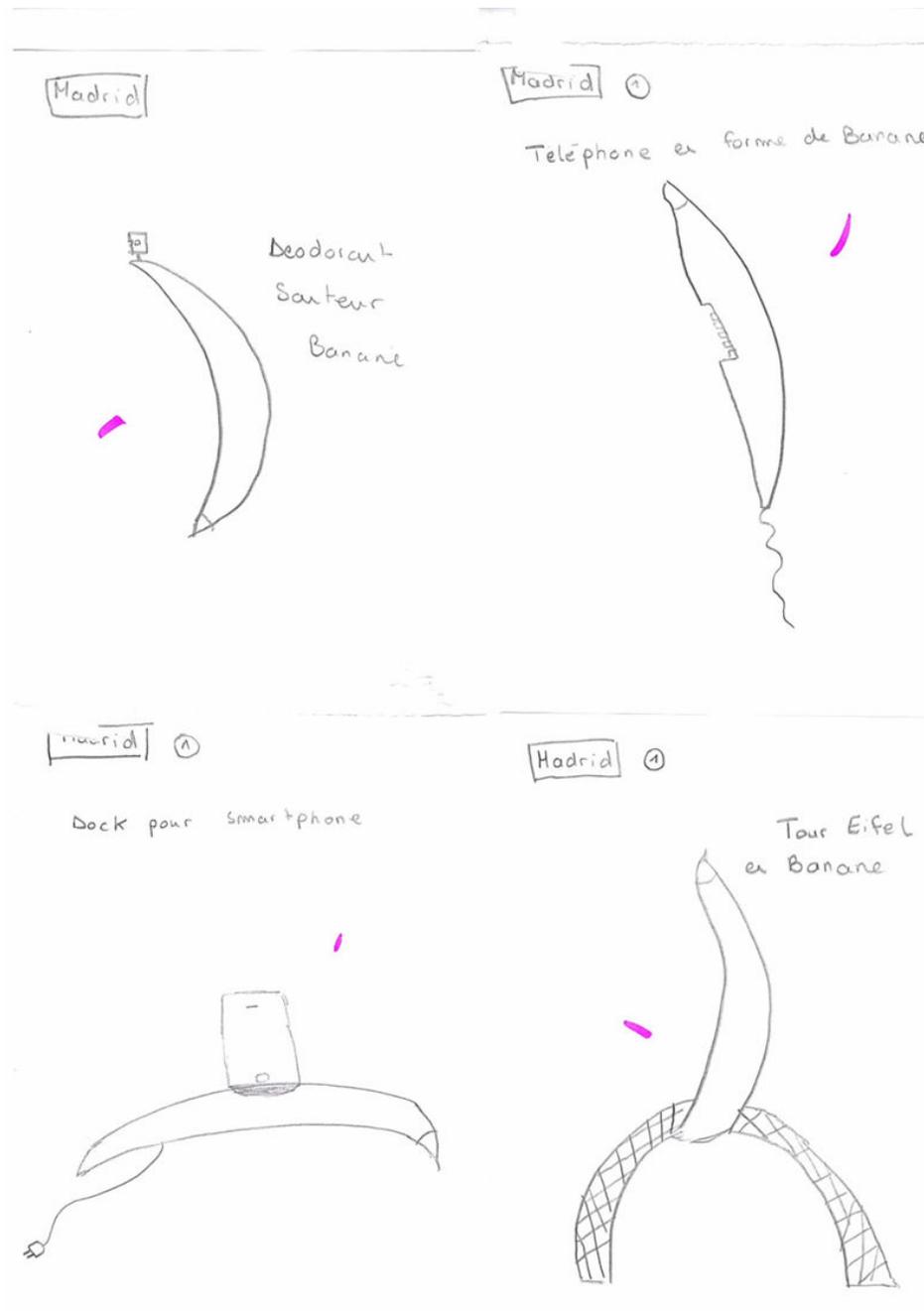


Figure 2.8 – Examples of convergent thinking graphical test high scores

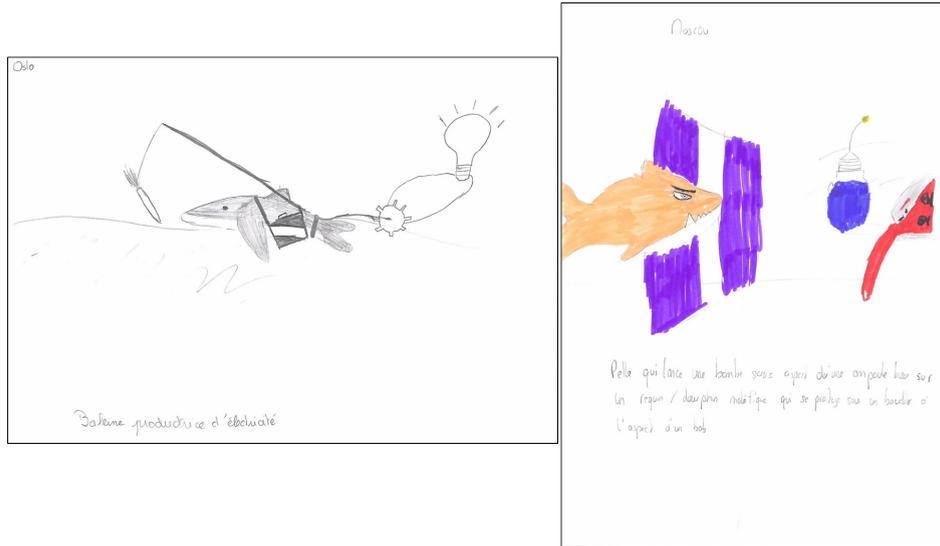
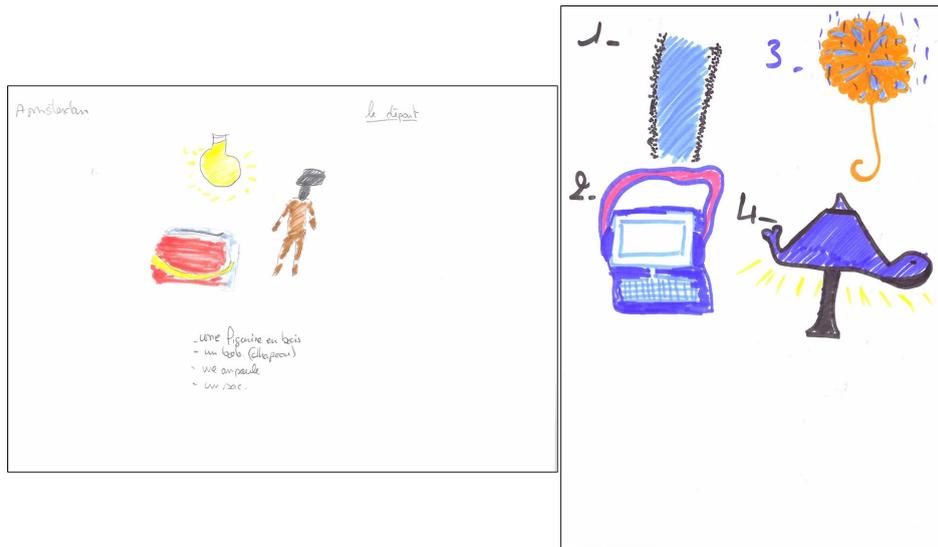


Figure 2.9 – Examples of convergent thinking graphical test medium scores



Figure 2.10 – Examples of convergent thinking graphical test low scores



A.6. Buttons task

Description: participants are randomly assigned to two different conditions, "performance" and "competition". We read the general instructions

Figure 2.11 – Buttons: comprehension questions (in French): "Performance treatment"

<p>1) Tu vas faire 3 séries et recevoir 3 récompenses. <input type="checkbox"/> Vrai <input type="checkbox"/> Faux {Réponse : faux, une seule série sera sélectionnée aléatoirement pour déterminer ta récompense. Tu dois faire de ton mieux les 3 fois, parce que tu ne sauras pas quelle fois aura été sélectionnée avant la fin de l'expérience.}</p>
<p>2) Ta récompense ne dépend que de ta performance. <input type="checkbox"/> Vrai <input type="checkbox"/> Faux {Réponse : vrai, pour chacune des série les points sont convertis en chocolat. Tu ne recevras que le montant de chocolat correspondant à la série sélectionnée aléatoirement à la fin.}</p>

Figure 2.12 – Buttons: comprehension questions (in French): "Competition treatment"

<p>1) Tu vas faire 3 séries et recevoir 3 récompenses. <input type="checkbox"/> Vrai <input type="checkbox"/> Faux {Réponse : faux, une seule série sera sélectionnée aléatoirement pour déterminer ta récompense. Tu dois faire de ton mieux les 3 fois, parce que tu ne sauras pas laquelle aura été sélectionnée avant la fin de l'expérience.}</p>
<p>2) Ta récompense ne dépend que de ta performance. <input type="checkbox"/> Vrai <input type="checkbox"/> Faux {Réponse : faux, Ta récompense dépend non seulement de ta performance, mais aussi de celle des autres. Tu ne recevras une récompense que si tu arrives dans les X premiers dans la série sélectionnée aléatoirement à la fin.}</p>

Figure 2.13 – Buttons - screen for series 1 and 3

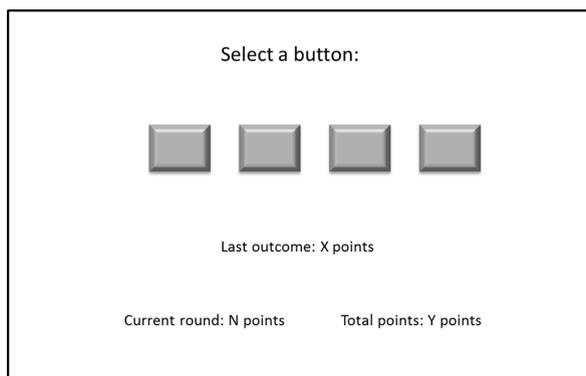
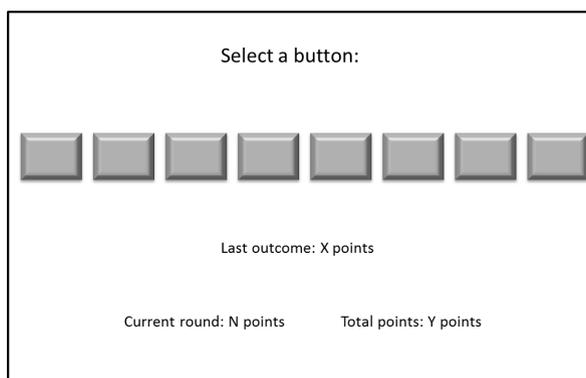


Figure 2.14 – Buttons - screen for series 2



After the task, two satisfaction questions:

1. Are you satisfied with this task? scale from 0 to 10
2. Are you happy with your performance? scale from 0 to 10

A.7. Typing task

Description: Participants must re-type code compounded by 5 letters (non-words, considering the entire keyboard). We present 150 possible codes (same for all participants in all sessions, same values and order), and they can type any of the codes in any order. After typing a correct code, it will be highlighted in the list of codes. We present on the screen the remaining time and the number of correct codes. Participants

perform the activity 4 times: two times alone and two times in pairs, only one of the activities is randomly selected to determine the reward.

Part II

School systems

INTRODUCTION TO PART II

The second part of this dissertation focuses on the experimental investigation of school systems, with two main objectives: the comparison of the efficiency of different school sorting mechanisms, and an evaluation of socioeconomic and gender biases caused by them. We compare *four mechanisms* for sorting students according to their abilities: self-selection of further studies with a single track (no-choice of track), self-selection of further studies with the choice of track (choice of track), screening by ability and early *numerus clausus* competition.

Chapter 3 focuses on the productive efficiency of schooling systems in sorting students by ability. We find evidence that the inefficiency of the system derives from two main reasons: (i) if students have an imperfect knowledge of their own ability, and/or if there is a lack of discrimination between educational tracks, they are inclined to opt for the more difficult track and fail; (ii) the higher is the wage premium for tertiary education the higher is the ex-ante expected utility, what increases the chances to try higher levels of education, but does not increase the probability of success, raising the ex-post inefficiency (higher level of failures, dropouts and regret).

The question studied in Chapter 4 is: *how do different school systems and school returns affect differently ability groups, genders, and social groups, thus causing substantial differences in socioeconomic and gender bias among developed countries and periods?* We observe that self-selection of educational track increases inequalities, but inequality decreases if the wage premium for tertiary education decreases. Also, random allocation of educational track is the only mechanism that is fair for gender differences.

Chapters 3 and 4 are based on the same experiment and dataset. This introduction puts in perspective elements that are common for both chapters.

1. A minimal school system

A *minimal school system*¹ is composed of three consecutive levels that are discontinuous, such that a student must succeed one level to start the next. Students invest time and effort in their education facing the risk of failure. This is crucial, since otherwise everybody could ask for diploma, revoking the signaling facet of school systems (Spence, 1973).

The primary education provides a general core knowledge for all students (*common syllabus*). It represents the longest period of education in students' life, lasting 9 years in general². After succeeding the common core syllabus, pupils must opt to start the secondary level of education (high school³) or to enter the job market. Individuals who decide for further studies must generally choose between the vocational and the general curricula (Figure II.1), two tracks of unequal difficulty providing more specialized knowledge in order to develop their occupational expertise (Lévy-Garboua et al., 2004). Normally a selection procedure is placed around this point to allow students with best records to opt for preferred schools and tracks⁴. If students had perfect knowledge of their own ability, the more able students would opt for the general and the less able for the vocational track.

The general curriculum is meant for more able students driving them to higher education (tertiary education), while the vocational track, that takes between 2 or 3

1. Even if school systems differ among countries and cultures, there are two main characteristics that are generally common in the architecture of a school system: (i) school systems have three levels of increasing difficulty, being the first compulsory in most developed countries, (ii) after a period of basic education, students choose between the vocational and the general curriculum. Concerning these two characteristics, what vary across school systems are the duration of the compulsory school, and the timing of the tracking in the vocational or general curriculum. The minimal school system described here reproduces the duration of compulsory education and the timing of the tracking of the French school system.

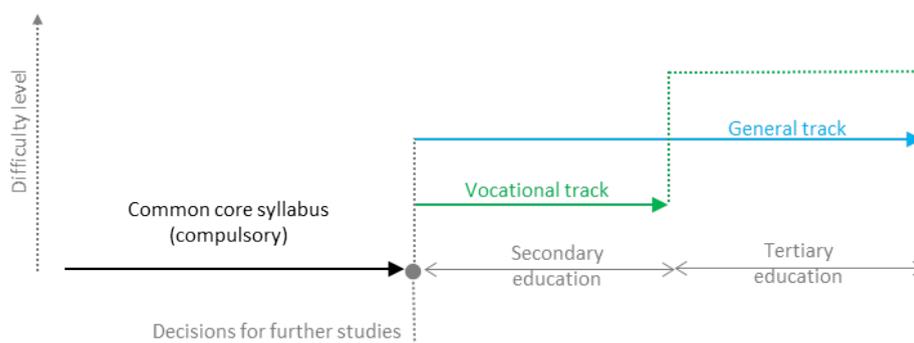
2. 5 years for the elementary school and 4 years for the middle school.

3. Note that high school is also compulsory in some countries, such as the United States of America.

4. There is a variety of situations, as the compulsory age and grade of bifurcation vary. We made a simplifying assumption by fixing the bifurcation at the compulsory age whereas it is sometimes before and sometimes after. The variety of situations is examined by Lassibille and Navarro-Gómez (2000).

years, provides an essentially technical (non-academic) training for students intending to work in manual or clerical jobs. It is important to observe that there is always a possibility of switching the track if admission requirements are met, in order to preserve the equality of chances. However, this option should not occur if students perfectly know their ability and/or institutions work perfectly in the selection. Anyway, students are less likely to attend college if they attend the vocational track (Arum and Shavit, 1995), mainly because they have a restricted curriculum, specially in maths and sciences (Gamoran, 1987). Thus, several studies suggest that tracking students by ability may promote inequality on educational outcomes (Wößmann, 2016). For instance, Hanushek and Wößmann (2006) show that it is not sure that early tracking promotes gains on students' average performance, but it surely increases the inequalities on educational outcomes.

Figure II.1 – School system: stylized design



Note: Decisions for further studies: (i) start secondary level or enter the job market?, and if starts the secondary level, (ii) vocational or general track?.

2. Methodology

There are several empirical limitations to study a given institutional context and/or make international comparisons as it is almost impossible to isolate the investigated effect maintaining everything else constant. The use of an experimental framework is a

good alternative to study educational institutions, even if it is not possible to capture all the elements of an educational system under a controlled environment. To address this research objective, we reproduce by means of an incentivized lab experiment the stylized educational system described by Figure II.1, where we can discriminate differences in curricula, differences in payoffs, the choice of the track and the performance level to qualify for a certain curriculum. The lab experiment allows us to measure variables like performance, ability, self-confidence and effort (time) that would be difficult to observe precisely in surveys. Experimental results allow an easy and valid comparison of the overall performance of a minimal school system under various sorting mechanisms, thus facilitating the identification of the efficient design that is, the educational output-maximizing design conditional on the ability distribution.

2.1. Experimental Design

The design takes its inspiration from the experimental school system described in Page, Lévy-Garboua and Montmarquette (2007). The latter is extended by introducing two tracks and mechanisms of allocation to them in some treatments. Participants perform a real-effort, rather long and difficult, task for which they get paid according to their degree of success. The task consists in solving anagrams ranked in three levels of increasing difficulty. Based on 15 years of study, taken from the actual minimal school system, our experiment is performed during a maximum of 15 rounds lasting no more than 8 minutes each. The three successive levels are designated respectively as the primary (level 1), the secondary (level 2), and the tertiary (level 3). Participants are successful at one level when they manage to decode 2/3 of the anagrams at this level. Succeeding the previous level is the condition to start a new level.

All participants start in the (compulsory) primary level, consisting of 9 rounds of low difficulty (*i.e.* 6 anagrams per round to be solved in no more than eight minutes). This level is long enough to let participants feel that a large effort and ability is required

of them to succeed at the optional upper levels. It does also let them ample time to learn the task and assess their own ability to perform it. Those who succeed this level can opt to stop after the primary level or to start the secondary level. Like educational decisions, the resolution to start the new level must be based on one's ability to perform the task. Those deciding to stop left the experiment with the money already earned, while those choosing start the new level can substantially increase their gains if they succeed to solve increasing numbers of anagrams under the same rules in two successive levels of three rounds each. However since there is a cost to start a new level, they lose part of their earnings and step out of the experiment if they fail to reach any of these levels.

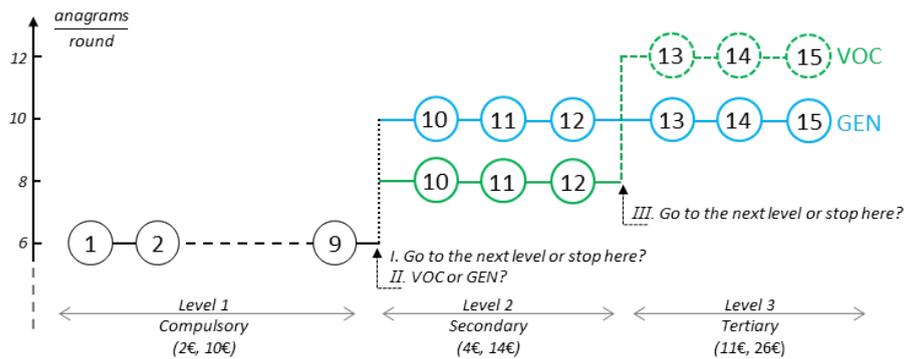
We manipulate the gradient of task difficulty after completion of the primary level (common syllabus) proposing two different educational curriculum: *(i)* in the *General* (GEN) condition, the difficulty jumps sharply at secondary level, but remains constant at tertiary level. General education is supposed to enhance the cognitive ability, i.e. the ability to learn faster, which facilitates the successful pursuit of higher studies. We transpose this characteristic of general education experimentally by letting the (experimental) tertiary level be no more difficult than the secondary level; *(ii)* in the *Vocational* (VOC) condition, the difficulty always rises from one level to the next, slowly first at secondary level, then sharply at tertiary level⁵. As on educational systems, an individual can reach the tertiary level after starting in the VOC track, but after an easier secondary level one faces a harder tertiary level if compared to individuals on the GEN track.

By the end of the experiment, the required number of anagrams is the same for the GEN and VOC conditions as school systems commonly aim to provide equal opportunities to all students to reach upper levels if they meet pedagogical requirements.

5. Terms used here (Vocational, General, primary level, secondary level and tertiary level) were not explicit to participants. During the experimental sessions we refer to Track I, Track II, level 1, level 2 and level 3.

However, the distribution of anagrams to be decoded differs for these two conditions. In the GEN condition, ten anagrams per round are proposed at the second and third levels, of which 20 anagrams at least must be decoded per level. In the VOC condition, eight anagrams per round are proposed at second level, and this rises to twelve anagrams at third level. Decoding sixteen anagrams in three rounds is required for second level; and decoding twenty-four anagrams in three rounds is required for third level. The required number of anagrams solved for passing the three levels is the same in these two conditions, describing the human capital's requirement for attaining the highest level of skill. This design and payoffs can be visualized in Figure II.2.

Figure II.2 – Experimental design



Notes:

Payoffs in parentheses : (fail, success and stop).

Decisions I, II and III are conditioned to success in the previous level.

Decision II depends on the treatment.

Treatments⁶

— **Random allocation of track (No-choice⁷)**: the No-choice treatment imposes the track to participants who succeed the compulsory level, it is composed by two sub-treatments:

— **GEN** track is imposed to participants who passed the primary level;

6. All participants are in the same condition in a given session.

7. The statistics reveal that no significant difference is noticeable in the results of the sub-treatments VOC and GEN. Thus, we aggregated these treatments in the No-choice condition.

- **VOC** track is imposed to participants who passed the primary level;
- **Voluntary allocation of track (Choice)**: a choice among the two conditions (VOC or GEN) is proposed to all participants who passed the primary level.
- **Voluntary allocation of track with low incentives (Choice-low)**: exactly as the Choice treatment, a choice among the two conditions (VOC or GEN) is proposed to participants who passed the primary level. The only difference to the previous treatment is that the payoff to reach the tertiary level is lower than the payoff for other treatments. In this treatment participants earn 18€ if complete the third level and 8€ if fail this level, instead of 26€ and 11€ for other treatments.
- **Constrained allocation of track (Screening⁸)**: like other treatments, the minimum requirement to clear the primary level is 36 out of 54 anagrams (2/3). If a participant solves between 36 and 44 anagrams in the first level, we impose the VOC track if she decides to start the secondary level. If she solves 45 or more anagrams in the first level, she has the opportunity to select which track (GEN or VOC) she prefers, like in the Choice treatment. The idea of the new threshold⁹ is to select the more able participants and give them opportunity to decide which track they would like to proceed, and impose to lower ability individuals the VOC track, with higher chances to succeed in the secondary level.

8. We have three sub-treatments in the Screening: (i) *full-information* of the screening procedure, (ii) *limited-information* on the screening: participants are instructed to solve as many anagrams as they can manage at the primary level. Solving more anagrams will render an advantage for the secondary level. No mention of screening procedure or the option of the choice was made to the individuals, (iii) *no-information*: no instructions were imparted to participants on the screening procedure. Individuals were only informed about the minimum number of anagrams required to be solved at the primary level. Relevant information about the secondary and tertiary levels was provided at the start of each level. As the overall results are non significant the three sub-treatments have been aggregated into a single version, the Screening treatment.

9. The threshold of 45 anagrams to select the more able participants is based on the performance of participants on the No-choice and Choice treatments, which sessions were performed before the screening treatment. High ability individuals solve on average 44.34 anagrams in the No-choice and Choice treatments.

- **Competitive allocation of track with rationing (on a first-done-first-served basis) (Race)**: a competition in the form of race is introduced in this treatment. The first eight participants (over 20) attaining 36 anagrams in the primary level are eligible for choice in the secondary level, while the VOC track is imposed for other participants that succeed the first level. It replicates the conditions of an early *numerus clausus*¹⁰.

Confidence Judgments Participants who pass the primary level and decide to continue the task are asked to state their subjective probability of success for the secondary and tertiary levels on a scale of 0 to 100 (Adams, 1957). The Adams’s (1957) scale that we use is convenient for quantitative analysis because it converts confidence into (almost) continuous subjective probabilities. It is required for consistency that the reported chances of success do not increase as the difficulty level increases. Answers cannot be validated as long as they remain inconsistent.

We did not directly incentivize beliefs because our primary aim was not to force subjects to make optimal forecasts of their chances of success but to have them report sincerely their true beliefs in their attempt to maximize their subjective expected utility. The true beliefs are those which dictate actual behavior following such prediction, and the latter was incentivized by the money gains based on subjects’ performance and decisions to start higher levels in the task. Considering that self-reports perform nicely while being much simpler and faster than incentive-compatible rules, use of the self-report seemed appropriate in this experiment¹¹.

10. We analyze the external validity of the Race treatment on [Appendices to Part II](#), section [A.1](#), page [135](#).

11. A longer discussion about incentives is displayed on [Confidence biases and learning among intuitive Bayesians](#) section [2.4](#), page [41](#).

2.2. Experimental sessions

For a total of 941 participants¹², we ran 53 experimental sessions at the BULCIRANO lab (Center for Interuniversity Research and Analysis on Organizations), Montreal (Canada), and at the LEEP (*Laboratoire d'Economie Expérimentale de Paris*), Centre d'économie de la Sorbonne. The difference between Paris and Montreal¹³ was observed to be insignificant. A show-up fee of 5€ in Paris and Can\$ 5 in Montreal was paid to the participants (from now on, all money amounts will be given in Euros). We recruited an equal proportion of women and men to favor gender comparison. About 80% of the participants are students.

At the start, instructions were read out and a hard copy of it was also provided individually. Participants answered six questions to test their full comprehension of the experiment. Information on gender, age, educational level and labor market status was required. The last question was a hypothetical choice between 5€ for sure and an ambiguous urn containing 100 balls of two colors (white and black) in unknown proportions. Ten Euros (10€) were to be earned if a black ball was drawn. Choice of the sure gain provided a rough but simple measure of risk aversion in the uncertainty context of the experiment.

3. Descriptive Statistics

Table II.1 presents the descriptive statistics by treatment. The results of statistical tests show that samples are homogeneous for all treatments. No significant difference

12. Our sample is composed by 78.4% of students in diverse domains (33% economics, 14% law, 11% management and business, 42% other). 56% of non-students are employed, 29% unemployed and 15% inactive). Age is the only difference between students (22.7) and non-students (33.5) that is statistically significant ($p = 0.0000$, two-sample Wilcoxon rank-sum (Mann-Whitney) test). The proportion of women, risk averse and the measure of task specific ability are not significant at 10% level. Main outcomes (payoff, level attained, total anagrams solved) and decisions (to start a new level or track selection) are not significant at 10% for students and non-students.

13. We have 8 sessions in Montreal and 45 in Paris.

at usual levels is observed among the samples' means for individual characteristics with the exception of Race participants, who differ on age, risk aversion, and ability when compared to other treatments.

Table II.1 – Descriptive statistics by treatment

	No-choice	Choice	Screening	Race	Choice-low
Women	0.48	0.51	0.49	0.48	0.50
Age	25.19 (7.85)	25.13 (7.49)	24.94 (8.58)	24.02 (7.50)	23.65 (3.61)
Risk Averse	0.57	0.51	0.59	0.47	0.51
Student	0.79	0.73	0.82	0.87	0.71
Ability	0.45 (0.25)	0.43 (0.25)	0.44 (0.25)	0.50 (0.24)	0.41 (0.24)
Observations	207	203	162	190	179

Notes: Standard deviation in parentheses. Statistical tests presented in Table II.2.

Table II.2 – Statistical tests for Table II.1

<i>p</i> – value	No-Choice			Choice		Screening	Choice-low
	Choice	Screening	Race	Screening	Race	Race	Choice
Women	.4905	.7428	.9658	.7276	.5342	.7837	.8151
Age	.5771	.5332	.0480	.1986	.0070	.1442	.7792
Risk averse	.2408	.6953	.0481	.1245	.4011	.0208	.9995
Student	.1673	.3998	.0296	.0295	.0006	.1801	.6848
Ability	.2269	.6376	.0694	.4929	.0033	.0208	.6011

Notes: Choice-low is only comparable to Choice. **Statistical tests:** Women, risk averse and student: two-sample test of proportions. Age and ability: two-sample Wilcoxon rank-sum (Mann-Whitney) test.

The (normalized) number of anagrams solved per minute in the first four rounds in the primary level is our estimated specific ability to solve anagrams¹⁴, since the main objective of the task is to solve two thirds of the proposed anagrams in a limited amount of time. We select the measure in the first four rounds because it is supposedly an exogenous and good measure of prior ability¹⁵ to solve anagrams. We postpone the discussion about differences on ability for Race individuals to Chapter 3, section 4.2.

14. We normalize the ability dividing values by 6, that is the highest number of anagrams solved per minute in our sample.

15. In fact this measure is not completely exogenous because participants decide how much time (t) they spend to solve each anagram, since when facing a difficult anagram they are able to decide if they devote more time to find the solution, or skip to the next anagram or to the next round.

A. Appendices to Part II

A.1. Race treatment

The objective of this treatment is to replicate the conditions of early *numerus clausus*, limiting the number of participants who may perform the task at the GEN track. We include the taste of competition and select the more able participants: the 8 participants reaching the threshold to succeed the compulsory level faster for a given experimental session. Time pressure is an important characteristic of educational evaluations, working fast to succeed an exam under time constraints is an important facet of more able students (Siegel, 1989).

We confirm that faster participants are the more able to perform our task in tables II.3 and II.4, it corroborates to the effectiveness of Race's mechanism of selection. We observe that those "potentially" selected in the Race ("first-eight") have statistically higher outcomes (lower rate of failure at secondary level, higher rate of success at tertiary level, higher payoffs and higher number of solved anagrams) if compared to those ("others") whom cleared the compulsory level and started the secondary level¹⁶, but would not be selected by the Race mechanism.

Results presented in these tables (II.3 and II.4) correspond to a simulation of Race's selection in the data collected for the No-choice and Choice treatments. We create two groups to make this comparison: the "first-eight" corresponding to the eight first clearing the compulsory level in a given session, and the "others" matching to the rest of participants in the same session. It is important to highlight that there is no mechanism of selection in the No-choice and Choice, and that participants were not aware about the importance of being fast to select the track for post-compulsory levels. Clearly, the ex-post performance of our sample shows that the more able are also those

16. We limit this analysis to the sample clearing the compulsory level and starting the secondary level. Differences are even higher when observing all sample.

solving anagrams faster.

Table II.3 – Level attained

	Fail level 2		Clear L3	
	No-choice	Choice	No-choice	Choice
"First-eight"	0.61	0.65	0.30	0.20
"Others"	0.89	0.94	0.09	0.04
<i>p-value</i>	0.0003	0.0001	0.0003	0.0067
<i>n</i>	139(83)	136(81)	139(83)	136(81)

n: all participants starting second level (first-eight). Wilcoxon signed-ranks

Table II.4 – Outcomes

	Payoff		Number of solved anagrams	
	No-choice	Choice	No-choice	Choice
"First-eight"	9.88	7.53	69.12	70.70
"Others"	5.87	4.69	55.98	54.54
<i>p-value</i>	0.0005	0.0001	0.0000	0.0000
<i>n</i>	139(83)	136(81)	139(83)	136(81)

n: all participants starting second level (first-eight). Wilcoxon signed-ranks

Another alternative to mimic school systems should be a rank of the maximum number of anagrams solved in the primary level, the maximum "educational" output. We did not opt for this treatment in our experiment because it would not be comparable to other treatments, since participants would be incited to solve all anagrams at compulsory level, expending much more time in this condition, potentially expending the 72 minutes available for the compulsory level¹⁷. Consequently, there are several experimental constraints to include this treatment in our experimental setting: (i) we could not pay more, for more time in the lab, because we must compare to other treatments; (ii) we could have an effect of fatigue and affect their performance on the following levels; (iii) if we have used this mechanism, we should have to wait for all people concluding the task to make the rank and some participants (the more able) should be waiting a lot of time without any activity in the lab.

17. In our experiment, the mean time spent in the primary level is 25 minutes ($SD = 12$).

A.2. Instructions in French

Instructions for the Choice-low treatment

INSTRUCTIONS

Vous participez à une expérience dans laquelle on vous demande de prendre des décisions. Chaque participant prend ses décisions individuellement devant son ordinateur. Au cours de l'expérience, vous pouvez gagner de l'argent. Le montant final de vos gains dépend de vos décisions. Vos gains vous seront versés intégralement à la fin de l'expérience.

L'expérience comporte 15 étapes : **à chaque étape, vous devez résoudre des anagrammes**. Plusieurs mots dont les lettres sont dans le désordre vous sont donnés. Exemple : « jrbnoou » apparaît à l'écran, vous devez trouver et écrire « bonjour ».

Veillez noter quatre précisions :

- [1] Il vous est demandé de trouver un mot précis, et non pas n'importe quel mot à partir des lettres. Exemple : il vous est demandé "balle" à partir de "ablel" = "label" sera refusé, même si c'est un mot bien orthographié, car il ne correspond pas au mot demandé. Ne vous étonnez donc pas si des mots existants sont refusés. En d'autres termes, il n'y a qu'une seule bonne réponse pour chaque anagramme.
- [2] Les mots solutions peuvent contenir des accents, mais les anagrammes et leur solution seront écrits sans accent : " eemm" = "meme".
- [3] Les solutions ne sont pas des verbes conjugués, et ne sont pas des mots accordés au féminin ou
Si la solution est un verbe, il est à l'infinitif (exemple "trouver"). Vous n'avez donc pas à chercher des solutions telles que "trouvas", "trouvait", "trouvées".
- [4] Les réponses doivent être écrites en "minuscule".

Les 15 étapes se découpent en trois niveaux. A chaque **niveau**, vous devez avoir en moyenne répondu à 2 anagrammes sur 3 pour passer au niveau suivant. La difficulté

	Branch I	Branch II
Étapes 10 à 12 (Niveau 2) :	8 anagrammes/étape	10 anagrammes/étape
Étapes 13 à 15 (Niveau 3) :	12 anagrammes/étape	10 anagrammes/étape

des anagrammes devient de plus en plus grande à chaque niveau, et le temps disponible pour les résoudre diminue proportionnellement.

Niveau 1 : De l'étape 1 à 9 Vous avez à chaque étape **6 anagrammes** à résoudre. Vous pouvez passer à l'étape suivante sans avoir résolu tous les anagrammes, il faut cependant que vous ayez résolu 36 anagrammes (sur un total de 54) à la fin de la neuvième étape pour franchir le niveau 1.

A l'issue de l'étape 9, si vous ne franchissez pas le niveau 1, le jeu s'arrête, vous gagnez **2€**. Si vous franchissez le niveau 1, le montant total de vos gains devient **10€**. Vous devez alors choisir soit d'arrêter, soit de passer au niveau 2.

Vous ne pouvez quitter l'expérience avant la fin de l'étape 9. Vous pouvez quitter l'expérience à partir de la fin de l'étape 9, dès que vous le souhaitez, en cliquant sur le bouton "quitter".

Après l'étape 9, vous devez vous acquitter d'un coût pour participer au niveau suivant si vous continuez. Ce coût est soustrait à vos gains antérieurs.

A ce stade, vous devez choisir entre deux options codées comme suit : "**Branch I**" et "**Branch II**".

Le nombre d'anagrammes à résoudre dépend de l'option choisie. Veuillez examiner attentivement le nombre d'anagrammes de chaque option avant de faire votre choix. En effet, votre choix est définitif et vous ne pouvez pas faire marche arrière.

Description des deux options :

Niveau 2 : De l'étape 10 à 12 Le coût pour participer au deuxième niveau est de **6€**.

Vous avez à chaque étape

$$\left\{ \begin{array}{l} 8 \text{ anagrammes par étape (Branch I)} \\ 10 \text{ anagrammes par étape (Branch II)} \end{array} \right.$$

à résoudre.

Vous pouvez passer à l'étape suivante sans avoir résolu tous les anagrammes, il faut cependant que vous ayez résolu:

$$\left\{ \begin{array}{l} 16 \text{ anagrammes sur un total de 24 (Branch I)} \\ 20 \text{ anagrammes sur un total de 30 (Branch II)} \end{array} \right.$$

à la fin de la douzième étape pour franchir le niveau 2.

À l'issue de l'étape 12, si vous ne franchissez pas le niveau 2, le jeu s'arrête, vous gagnez $10\text{€}-6\text{€}= 4\text{€}$. Si vous franchissez le niveau 2, le montant total de vos gains devient $20\text{€}-6\text{€}=14\text{€}$. Vous devez alors choisir soit d'arrêter, soit de passer au niveau 3.

Niveau 3 : De l'étape 13 à 15 Le coût pour participer au troisième niveau est de **6€**.

Vous avez à chaque étape:

$$\left\{ \begin{array}{l} 12 \text{ anagrammes par étape (Branch I)} \\ 10 \text{ anagrammes par étape (Branch II)} \end{array} \right.$$

à résoudre.

Vous pouvez franchir le Niveau 3 sans avoir résolu tous les anagrammes, il faut cependant que vous ayez résolu:

$$\left\{ \begin{array}{l} 24 \text{ anagrammes sur un total de 36 (Branch I)} \\ 20 \text{ anagrammes sur un total de 30 (Branch II)} \end{array} \right.$$

à la fin de la quinzième étape pour franchir le niveau 3.

À l'issue de l'étape 15, si vous ne franchissez pas le niveau 3, le jeu s'arrête, vous gagnez $14\text{€}-6\text{€}= 8\text{€}$. Si vous franchissez le niveau 3, le montant total de vos gains

devient $24\text{€}-6\text{€}=18\text{€}$.

Temps limite A chaque étape, de l'étape 1 à 15, le temps que vous consacrez à résoudre les anagrammes ne peut excéder **8 minutes**. Si vous dépassez cette limite lors d'une étape, vous êtes directement invité à passer à l'étape suivante. Vous êtes informé lorsqu'il vous reste trois minutes. Vous ne pouvez donc dépasser 8 minutes pour résoudre :

- Les 6 anagrammes à chaque étape (de 1 à 9) du niveau 1 ;

$$\left\{ \begin{array}{l} \text{Les 8 anagrammes (Branch I)} \\ \text{Les 10 anagrammes (Branch II)} \end{array} \right.$$

à chaque étape (de 10 à 12) du niveau 2 ;

$$\left\{ \begin{array}{l} \text{Les 12 anagrammes (Branch I)} \\ \text{Les 10 anagrammes (Branch II)} \end{array} \right.$$

à chaque étape (de 13 à 15) du niveau 3.

Possibilité d'interrompre l'épreuve Vous ne pouvez interrompre l'épreuve lors des 9 premières étapes. À partir de la fin de la neuvième étape, vous pouvez décider d'arrêter à tout moment. Lorsque vous arrêtez, vos gains sont calculés en fonction de votre dernière épreuve atteinte, selon les modalités précisées précédemment.

Renseignements complémentaires Avant de débiter la session expérimentale, nous allons vous poser quelques questions de compréhension sur ces instructions. Dès que vous aurez tous répondu correctement à toutes les questions, nous vous prions de bien vouloir nous fournir des renseignements concernant votre âge, sexe, niveau et discipline d'études, votre situation par rapport au marché du travail, et si vous avez déjà participé à une expérience. Ces informations resteront anonymes. Vous devez répondre aussi un petit questionnaire à la fin de la session. L'expérience pourra alors débiter.

Merci de lever la main si vous avez des questions concernant ces instructions. Nous y répondrons avant de commencer. Si vous avez par la suite des questions, nous viendrons vous répondre personnellement.

Remarque Importante : Forfait de participation : 5€

CHAPTER 3

AN EXPERIMENTAL COMPARISON OF THE EFFICIENCY OF SCHOOL SYSTEMS

This chapter is a joint work with Louis Lévy-Garboua and Muniza Askari.

1. Introduction

The economic literature on education institutions and systems arose from the assumption of an informational asymmetry between workers and employers. In theories of signaling (Spence, 1973), filtering (Arrow, 1973), screening (Stiglitz, 1975) and sorting (Weiss, 1983), schools and universities provide certified information on students' abilities to employers and save them the cost of testing the abilities of job candidates. This was a major breakthrough after human capital theory (Becker, 1964) which assumed perfect information and abstracted from the school system.

Human capital theory is a theory of learning which only requires firms transmitting knowledge, maybe but not necessarily specialized. Besides, the Beckerian analysis describes a continuous investment which leaves no room for classes, thresholds, repetitions, dropouts, as well as educational tracks. The analysis of educational systems starts where the conventional human capital framework ends.

The higher the wage premium for tertiary education, the higher the demand for education (Fredriksson, 1997). However, as there is risk of failure at school, educational decisions must be grounded not only on economic incentives but also on chances of success for further studies. If on one hand the economic incentives are more explicit, on the other the probability of future success is hard to estimate. Increasing the demand for education may push low-ability individuals to try higher levels of education enhancing the ex-post inefficiency of the system, even if decisions are ex-ante efficient since individuals maximize their expected utility.

The chances of success at each level of education depend on individual characteristics, in particular cognitive and non-cognitive skills (Borghans et al., 2008). Of course IQ is an important predictor of success at school, but it is not a guarantee of future success if the individual is not motivated by her studies. Achievement motivation is also important for academic success (Busato et al., 2000) and has a direct impact on

the students' perception that success depends on one's effort (Ames and Archer, 1988). Even if effort (and motivation) explains why students with different ability levels can reach the same educational outcome, the economic literature focuses mostly on ability in order to explain academic success mainly because it is hard to measure and compare the effort of students.

More recently Heckman and Rubinstein (2001) put emphasis on the importance of non cognitive skills that are important to academic success, such as: persistence, self-esteem, risk tolerance, optimism and time preferences. For instance, Castillo et al. (2011) found that more patient children, those who are concerned with future consequences of their behavior, have a more favorable outlook for school performance. We observe in this study two noncognitive skills that may be responsible for large behavioral disparities between individuals: *self-confidence and motivation*. Self-confidence is measured by the reported chances of own successful performance in a real-effort task (solving anagrams) whereas motivation is captured by the time and effort attended to this task in comparison with the maximum allowance.

This study focuses on the productive efficiency of school systems in sorting students by ability and supporting educational decisions. School systems' efficiency implies an economic state in which every resource is optimally allocated, serving each agent in the best way. In other words, an efficient school system maximizes students' educational expected outcomes. Equivalently, an efficient school system should minimize expected regret¹, which is associated to failures and dropouts. If students knew their own ability, it would be efficient to let them self-select the level and track maximizing their expected utility because they would have a precise and accurate knowledge of their

1. In expected utility (EU) theory, it is known that for all A, B : $EU(A) - EU(B) = EOL(B) - EOL(A)$, with $EOL(A)$ designating the expected opportunity loss of A with respect to B (see Raiffa, 1968, for instance). If $EOL(A)$ is the measure of expected regret of choosing A and foregoing B . So, maximize EU is equivalent to minimize expected regret. Although the two programs are equivalent by duality, it is more common to speak of EU maximization. In the context of education policies, the value of speaking of regret is because ex-post regret feeds relative frustrations and political discontent. However, the two objectives: maximize EU and minimize expected regret yield the same conclusions.

chances of success. However, maximizing EU does not guarantee success nor does it guarantee the estimated chances of success if students have an imperfect knowledge of their ability. Students who failed and suffered an opportunity loss will *regret* their choice ex-post and ask for political redistribution. While political platforms aim at reducing failures -viewed as wastage, or ex-post inefficiency-, students pursue their own objective of maximizing EU with an imperfect knowledge of their chances of success. Thus, education is the life domain with highest potential of regret in contemporary society (Roese and Summerville, 2005).

The question is, then, to *design an education system that minimizes (ex-post) expected regret when students maximize their expected utility with an imperfect knowledge of their ability*. We compare *four mechanisms* for sorting students according to their abilities: self-selection of further studies with a single track (no-choice of track), self-selection of further studies with the choice of track, screening by ability and early *numerus clausus* competition. We find evidences that the inefficiency of the system derives from two main reasons: (i) if students have an imperfect knowledge of their own ability, and/or if there is a lack of discrimination between educational tracks, they are inclined to opt for the more difficult track and fail; (ii) the higher is the wage premium for tertiary education the higher is the ex-ante expected utility, what increases the chances to try higher levels of education, but does not increase the probability of success, raising the ex-post inefficiency (higher level of failures, dropouts and regret).

This paper is organized as follows: section 3 analyzes the efficiency of different school systems. We find that Race and Choice are not efficient for different reasons: the problem of Race is the high level of failure at compulsory level, whereas the problem of Choice is the remarkable rate of failure at secondary level. Section 4 aims at explaining the cause of inefficiencies of Choice and Race. Section 5 proposes a model of effort based on ability and subjective probability of success. Section 6 evaluates the impact of incentives on the efficiency of educational systems showing that higher returns to

tertiary education increase the amount of regret. Finally, section 7 tests the model presented in section 5 on our experimental data. Conclusions are presented in section 8.

2. Experimental design, experimental sessions and descriptive statistics

The procedure of the experiment and details of the experimental sessions are described in the introduction to Part II (page 127). Descriptive statistics of our 941 participants are also provided in the introduction to Part II (page 133).

3. Efficiency of school systems

3.1. Performance in the task

We have three main criteria to evaluate participants' performance in the task: payoff and (educational) output (E)², and the maximum level attained (l). The higher are payoffs, outputs and rate of success at post-compulsory levels, and the lower are the dropout rates (failures), the higher is the productive efficiency of the treatment. We don't analyze the Choice-low treatment in this section since it is only comparable to the Choice treatment. We postpone the analysis of the Choice-low to Section 6.

Result 1: No-choice and Screening are the best treatments, Race is the worst. Taking No-choice as reference, the value of education produced diminishes by 19.8% in the Race treatment, the educational output diminishes by 11.4%, and the ratio of individuals succeeding the highest level of the task decreases by 42.9% (Table 3.1). The level attained in the Race is first-order stochastically dominated by No-choice³.

2. Measured by the number of solved anagrams.

3. See Figure 3.5 in the Appendix for more information

Race is by all means the worst treatment because participants in the Race present the worst result for these three measures: lower payoffs⁴, lower outputs, and lower chances to reach higher levels in the task.

Table 3.1 – Efficiency of treatments

	No-choice	Choice	Screening	Race
Payoffs (€)	9.00 (7.78)	7.80 (7.09)	8.18 (7.18)	7.22 (6.72)
Output (E)				
Total anagrams solved	54.65 (19.88)	54.32 (20.93)	56.38 (19.57)	48.42 (18.19)
Anagrams solved at compulsory level	40.59 (7.14)	41.03 (9.27)	42.64 (8.14)	37.22 (7.59)
Maximum level attained (I)				
Dropouts (fail compulsory level)	0.13	0.13	0.14	0.23
Compulsory level	0.63	0.70	0.64	0.59
Post-compulsory levels	0.24	0.17	0.22	0.18

Notes: **Standard deviation** in parentheses. **Statistical tests** are reported in Table 3.2. For more information, output and maximum level attained are reported separately for the three levels in the Appendix (Table 3.12).

Table 3.2 – Statistical tests for Table 3.1

<i>p</i> – value	No-Choice			Choice		Screening
	Choice	Screening	Race	Screening	Race	Race
Payoffs (€)	0.0948	0.3428	0.0049	0.5032	0.1633	0.0572
Output (E)						
Total anagrams solved	0.7672	0.1449	0.0032	0.2090	0.0016	0.0000
Anagrams solved at compulsory level	0.2271	0.0007	0.0000	0.0671	0.0000	0.0000
Maximum level attained (I)						
Dropouts (fail compulsory level)	0.8235	0.6296	0.0075	0.7937	0.0145	0.0316
Compulsory level	0.0491	0.5523	0.1074	0.1796	0.7713	0.3113
Post-compulsory levels	0.2667	0.3068	0.0623	0.9456	0.4183	0.3878

Statistical tests: Payoffs: two-sample Wilcoxon rank-sum (Mann-Whitney) test. Total anagrams solved: two-sample t-test with equal variances. Level attained and decisions: two-sample test of proportions.

The performance in the No-choice and in the Screening are very much alike, and better than Choice. No-choice participants present on average higher payoffs and higher ratio of participants reaching post-compulsory levels, while the Screening promotes the highest number of anagrams solved, not significantly different at usual levels when compared to No-choice and Choice.

4. Payoffs for Race are lower than Choice but not significant at 10% level.

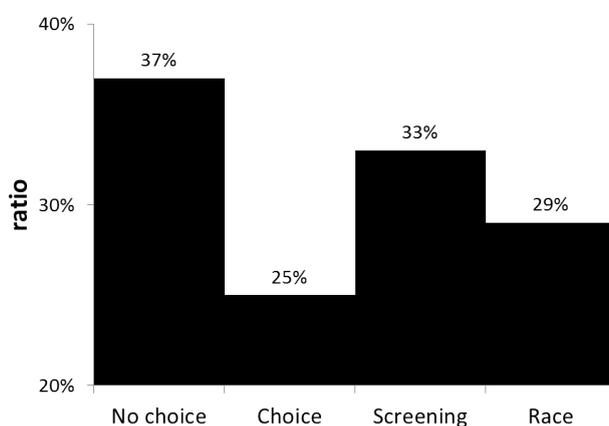
3.2. The effect of selection by ability

Screening and Race treatments propose different mechanisms to select participants by ability and give to the *selected* the opportunity to choose the preferred track to be performed at the post-compulsory levels. These mechanisms produce opposite effects. Table 3.1 shows that Screening generates the highest output and Race the lowest output at compulsory level. The threshold to enable the track selection in the Screening leads to 14.6% more solved anagrams than in the Race treatment. Moreover the dropout rate at the compulsory level is remarkable in the Race, 64.3% higher than in the Screening⁵. The question that arises is: *why do participants produce less and fail more in compulsory level in the Race treatment?*

3.3. The effect of self-selection of educational track

Result 2: Self-selection of track decreases efficiency at post-compulsory levels. In the Choice condition, the rate of success in the secondary level is 12% lower ($p = 0.0363$) relative to No-choice (Figure 3.1).

Figure 3.1 – Success at secondary level, conditional on starting this level.

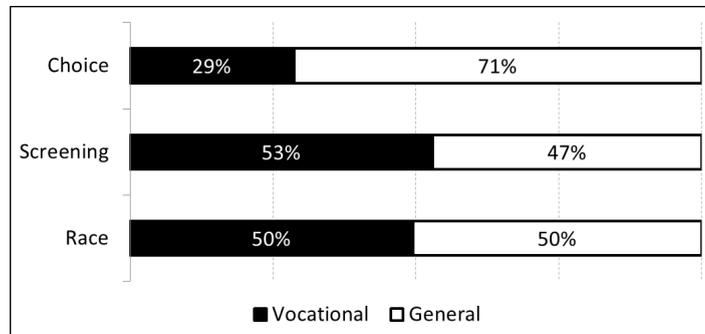


5. Dropout rates at compulsory level are alike for No-choice, Choice and Screening.

The efficiency of school systems, as replicated in our experiment, is directly connected to the three decisions that participants face along the task: decisions to start secondary and tertiary levels if they succeed the previous level, and selection of track (depending on treatment). Since there is no difference among treatments in the proportion of participants who succeed compulsory level and decide to start the secondary level⁶, the self-selection of preferred track may explain the inefficiency of Choice.

Figure 3.2 shows the distribution of participants among tracks: considering treatments in which participants could select the preferred track, Choice presents the highest ratio of participants in the general track (71%), the one with lowest chances of success at the secondary level (Table 3.3). The question that arises is: *why do many participants select the most difficult track when they have a choice, then fail and regret later?*

Figure 3.2 – Distribution of participants to post-compulsory education.



Result 3: Self-selection of educational track is costly. When analyzing the performance of Choice participants that selected GEN in Figure 3.3, only 23% of them solved 20⁷ or more anagrams and succeeded the secondary level. Note that 39% of this sample solve more than 15 anagrams but less than 20, which is not sufficient in the GEN but is enough to succeed in the VOC track where the threshold to clear the

6. Decision to start secondary level is reported in Appendix (Table 3.12).

7. In the GEN track participants had to solve 2/3 of 30 anagrams to pass the level.

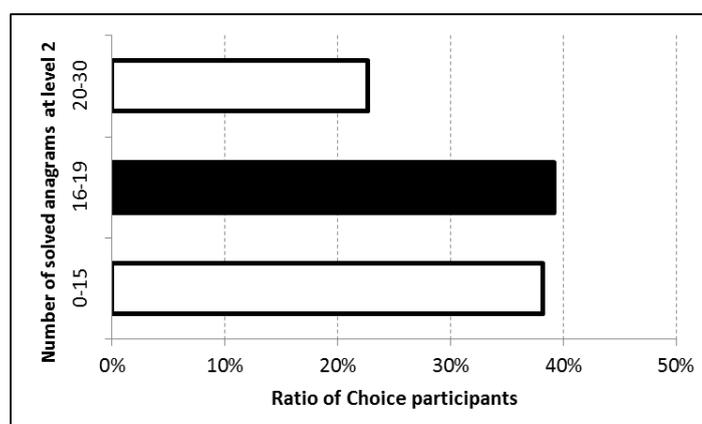
Table 3.3 – Chances of success for secondary level

	VOC	GEN	<i>sig.</i>
Chances of success (ex-post)	37%	26%	***

Sample: participants that started the secondary level. Chances of success (ex-post) are conditional on success at primary level and decision to start the secondary level. We report mean values for the aggregation of No-choice, Choice, Screening and Race treatments. **Statistical test:** two-sample Wilcoxon rank-sum (Mann-Whitney) test. Significance level (**sig.**): * 10%; ** 5%; ***1%; no symbol means that it is not significant at 10% level.

secondary level is 16 anagrams. It means that these participants make the wrong choice because they could have a higher payoff if they select the VOC track.

Figure 3.3 – Choice treatment: number of anagrams solved at secondary level in the GEN



Note: the number of anagrams solved at secondary level is conditional upon success in compulsory level and starting secondary level.

This single decision decreases the total wealth generated in the Choice treatment by 24%. To compute the *cost* of track self-selection we must assume that these participants could have the same performance if they selected the VOC track and that they should stop after the secondary level. In the Choice, we have 38 participants that selected GEN and failed, but could succeed VOC. The difference of payoffs for success and failure in the secondary level is 10€, so participants earned 380€ less in the experiment (the

total payment for the 203 participants in the Choice is 1584€)⁸.

The effect of the constrained choice of track in the Screening and Race treatments decreases the regret caused by this decision since those who choose GEN in these treatments have higher ability on average. Applying the same reasoning used in the Choice, we observe that the cost of track selection drops to 13.5% in the Screening and 12.4% in the Race⁹. Thus, when considering all treatments¹⁰, the effect of self-selection of track represents a loss of 12%¹¹ in the wealth generated for the whole experiment.

4. Why choice and early competition fail

The comparison of treatments in Section 3 points out the inefficiency of Choice and Race in comparison to No-choice (and Screening). Table 3.4 summarizes the main differences among treatments observed in the previous section.

Table 3.4 – Comparison to No-choice outcomes

	Choice	Screening	Race
Payoff	<i>lower*</i>	<i>ns</i>	<i>lower***</i>
Number of solved anagrams	<i>ns</i>	<i>ns</i>	<i>lower***</i>
Dropout rate (compulsory level)	<i>ns</i>	<i>ns</i>	<i>higher***</i>
Dropout rate (secondary level)	<i>higher**</i>	<i>ns</i>	<i>ns</i>
Attained tertiary level	<i>ns</i>	<i>ns</i>	<i>lower*</i>

Legend: *lower*: the outcome is lower than in the No-choice treatment, *higher*: the outcome is higher than in the No-choice treatment, and *ns*: the outcome is not statistically different at 10% level. **Statistical tests:** Payoffs: two-sample Wilcoxon rank-sum (Mann-Whitney) test. Total anagrams solved: two-sample t-test with equal variances. Level attained and decisions: two-sample test of proportions. **Significance level:** * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, *ns* non significant at 10% level.

We then need a theoretical investigation in order to clarify the two main questions that emerges in the previous section, which may explain the inefficiency observed in

8. If we consider that these participants could start and fail the tertiary level the welfare cost is 16.8% of the total income generated by this treatment.

9. There is no regret in the No-choice condition since participants don't select their track under this condition.

10. No-choice, Choice, Screening and Race

11. 8.3% if we consider that participants could start and fail tertiary level.

the Choice and Race:

- [1] *Choice paradox*: why do many participants select the most difficult track when they have a choice, then fail and regret later?
- [2] *Race paradox*: why do participants produce less and fail more in the compulsory level in the Race?

We propose first an intuitive -non formal- argument to answer these questions before developing a formal theory of educational choices in the next section.

4.1. Why choice of education fails: maximizing expected utility or minimizing regret?

The *choice paradox* arises from the massive selection of the GEN track followed by the highest rate of failure at the secondary level among all treatments observed in the Choice treatment. In fact, the dropout at secondary level is important for all treatments (Figure 3.1), so regret is not restricted to the selection of track, but extends to the decision of starting the secondary level with the risk of incurring an opportunity loss in case of failure. Thus, we extend the *choice paradox* to all educational decisions.

For both decisions, participants have a choice between low prospects with a high probability or high prospects with a low probability. Considering first the decision to start the secondary level after successful completion of the primary level, the decision to start the secondary level should be taken if the expected utility of continuing is higher than the sure utility of stopping. In order to maximize their expected utility, students must compare the expected utilities of the three available options ¹²:

- (i) Stopping after compulsory level:

$$EU_1 = u(y_1)$$

12. In the simple exposition of this section, we neglect the cost of effort. The latter is reintroduced in section 5.

(ii) Seeking to achieve a secondary education only:

$$EU_2 = (1 - q_2).u(y_{20}) + q_2.u(y_2)$$

(iii) Seeking to achieve a tertiary education:

$$EU_3 = (1 - q_2).u(y_{20}) + q_2.[q_{3|2}.u(y_3) + (1 - q_{3|2}).u(y_{30})]$$

where: q_2 is the subjective probability of success for secondary level, $q_{3|2}$ chances of success at tertiary level conditional upon success in the previous level, y_1 , y_2 and y_3 are payoffs for success at each level, y_{20} and y_{30} payoffs for failure at levels 2 and 3.

In our experiment, we observe: $q_2 = 0.31$ and $q_{3|2} = 0.72$ if we aggregate all treatments with same incentives. Hence, it is rational for well-calibrated risk-neutral subjects to seek a tertiary education since the expected value of the third option is equal to 10.92€, which is higher than 10€, the sure payoff of stopping after primary level, and than 7.10€, the expected value of seeking a secondary education only. This simple calculation illustrates the choice paradox: the best choice ex-ante may be the one that raises the highest expected regret ex-post. This paradox may occur in any intertemporal decision like lifetime investments, educational choices and career choice.

This kind of problem is commonly discussed in the literature on relative frustration and reference groups (for instance [Boudon \(1977\)](#)), and nicely illustrated by the famous result of *The American Soldier*. [Stouffer et al. \(1949\)](#) show that even if promotion is rare for police members, they feel satisfied with the system of promotion which governs them. Conversely, promotions for aviators is more frequent, but they feel dissatisfied with the system of promotion. Everything happens as if higher opportunities for promotion caused lower job satisfaction.

The same argument applies to the massive selection of the general track in the Choice treatment. The majority of students seeks a tertiary education and opts for

the general track because it yields a higher EU than the vocational track for students who seek a tertiary education¹³. Unfortunately, the latter is more difficult than the vocational track at secondary level so that many of them fail and will regret having dismissed this option.

Screening as a solution to regret. Screening is ex-post efficient because it aims at reducing failures and dropouts at secondary level preventively. The introduction of a higher threshold to enable the track selection narrows this decision for the more able, forcing the easiest track (VOC) for the less able. Thus, screening forces students to maximize EU under the constraint of low failure probability in the general track. A constraint for track selection is added by institutions in order to minimize expected regret if one does not know her own ability.

$$\begin{aligned} & \max EU \\ & \text{s.t: } \bar{q} \geq \bar{q}_G \text{ if choice of general track} \end{aligned}$$

where \bar{q} are the objective chances of success.

4.2. Why early competition fails: winning in the short run or in the long run?

During the Race, participants compete for selection, *i.e.* obtaining the right to choose their preferred track. Restricting the analysis to participants who would pass the first level, s.t. the (educational) output $E \geq 36$, the selection is made on the basis of the minimum observed time (t) for completing 36 anagrams.

Equivalently, participants are selected on the basis of the maximum educational

13. For more information, the expected utility for decisions to start secondary level and track selection is presented in the Appendix, section A.2, page 179.

productivity \bar{k} . Then, subjects maximize \bar{k} by allocating less time (and effort) to difficult anagrams in the first nine rounds of the task: $\bar{k}.t = \sum_{i=1}^9 \bar{k}_i \times t_i$ (choice of t_i). However, minimizing the time spent on solving anagrams during the first level is not efficient as it will not generally maximize EU. Equation 3.1 shows that Race participants minimize the time to reach 36 anagrams in the compulsory level.

$$\begin{aligned} V_{\text{Race}} &\equiv \min_{t_1, \dots, t_9} t \equiv \sum_{i=1}^9 t_i & (3.1) \\ \text{s.t: } &\sum_{i=1}^9 \bar{k}_i \times t_i = 36 \end{aligned}$$

Table 3.5 confirms that the competition in the Race induces participants to reach the goal for the primary level faster than other participants in the same experimental session. The time spent in the compulsory level is significantly lower in the Race compared to other treatments. The same table suggests that Race participants are faster because they devote less effort to solve hard anagrams. It is important to highlight that participants that participants don't need to win the Race in order to achieve the primary goal of the task, that is solving 2/3 of anagrams to succeed the level. Clearing the level and winning the Race are different and independent goals, and the first should dominate the later.

Table 3.5 – Efficiency of treatments

	No-choice	Choice	Screening	Race
Effort (e)				
Task specific time (t_6)	14.15 (0.55)	15.25 (0.53)	14.68 (0.56)	11.25 (0.47)
Rate of easy anagrams solved	91%	91%	92%	88%
Rate of hard anagrams solved	67%	71%	69%	58%

Notes: **Task specific time** (t_6) corresponds to the time spent in the six first rounds of the primary level, presented in minutes. **Rate of easy (hard) anagrams solved** considers the 18 anagrams with higher (lower) rate of success in the No-choice treatment among the 36 first anagrams in level 1. **Standard deviation** in parentheses. **Statistical tests** are reported in Table 3.6.

Table 3.6 – Statistical tests for Table 3.5

<i>p</i> – value	No-Choice			Choice		Screening
	Choice	Screening	Race	Screening	Race	Race
Effort (<i>e</i>)						
Task specific time (t_6)	0.1480	0.4926	0.0001	0.4611	0.0000	0.0000
Rate of easy anagrams solved	0.6133	0.3402	0.0148	0.2079	0.1007	0.0022
Rate of hard anagrams solved	0.0720	0.3549	0.0000	0.3898	0.0000	0.0000

Statistical tests: two-sample t-test with equal variances.

Result 4: Competition for self-selection of educational tracks causes regret. The competition for self-selection of track decreases the total wealth generated in the Race treatment. Considering that instead of taking care of relative performance, participants should be concerned with the main goal of the task -that is solving 36 anagrams at compulsory level-, the rate of success at compulsory level in the Race should be the same as in other treatments. Taking No-choice as reference, the rate of early dropouts should decrease by 10%, that is 18 participants who failed but could succeed the compulsory level. Considering the differences in payoff for success¹⁴ and failure at compulsory level (8 €), the total wealth generated in the Race condition should be increased by 144 €¹⁵, that is 11.1%.

5. Confidence, motivation and performance

In this section we introduce motivation -measured by effort- and other behavioral elements to develop a formal theory for educational decisions.

5.1. Definitions

The educational output (E), or human investment, is measured by the number of anagrams solved in our task. Assuming that effort (e) and educational output (E) are (potentially) measurable, the individual "ability" (k) is commonly defined by:

14. The cost of competition decreases to 2.8% if we consider that participants could succeed primary level, start secondary and fail.

15. The total payoff in the Race treatment is 1,293€.

$$E = k.e$$

However, we assume that the difficulty of the task and the individual performance have a random component so that the true ability can at best be approximated but remains imperfectly known. In fact, the exact relation between the educational output and effort is stochastic:

$$E = k.e + \epsilon \tag{3.2}$$

where ϵ is a random error term of zero mean.

Effort corresponds to the efficient time in the task. Participants allocate their task-specific time (t) between efficient time or effort (e) and *slack time* given to inattention, rest or leisure during the experiment. The *efficient time ratio* ($0 \leq \theta \leq 1$) defines the average intensity of effort, describing how task-specific time converts into "efficiency units". Whereas effort is painful and bears a utility cost, slack time is painless and provides direct utility in proportion of time spent under this condition.

$$e = \theta.t, \text{ with: } 0 \leq \theta \leq 1 \tag{3.3}$$

Combining equations 3.2 and 3.3:

$$\begin{aligned} E &\equiv (k.\theta).t \\ &= \bar{k}.t \end{aligned} \tag{3.4}$$

Where \bar{k} is the estimated ability corresponding to the productivity (speed) of solving anagrams - i.e. number of anagrams solved per minute.

The true probability of success (\bar{q}) in the experiment is a function of the number of solved anagrams:

$$\begin{aligned}
 \bar{q} &= Prob(k.e + \epsilon \geq \bar{E}) \\
 &= Prob(\epsilon \geq \bar{E} - k.e) \\
 &= 1 - F(\bar{E} - k.e) \\
 &= S(\bar{E} - k.e) \\
 &= \rho(k.e) , \text{ with: } \rho' > 0, \rho'' < 0
 \end{aligned} \tag{3.5}$$

where \bar{E} is the number of anagrams required for passing and $F(S)$ is the (de)cumulative distribution function.

5.2. Confidence

If they pass the first and second levels of the experiment, participants must decide to start the second and third levels, respectively, or to stop and stay with the payoff of the current level. As participants do not know perfectly their objective chances of success, this decision is based in the subjective chances of future success (q), that we call confidence. Realized output (E) at one level ($l = 1, 2$) enables one to forecast her probability of success at next level(s), but it does not guarantee an unbiased estimation of confidence since people commonly underestimate their ability to perform an easy task and overestimate their ability to perform a difficult task (this is the "hard-easy effect" revealed by [Lichtenstein and Fischhoff \(1977\)](#)). We borrow from Chapter 1 (Section 5) the analysis of the confidence (q) of participants on their future success, which predicts the hard-easy effect and other behavioral anomalies of confidence. Chapter 1

(Section 5) describes the systematic miscalibration of the probability of success (\bar{q}) by the following relation:

$$\begin{aligned} q &= \mu \bar{q} + (1 - \mu)D \\ q &= \mu \rho(ke) + (1 - \mu)D \end{aligned} \tag{3.6}$$

Confidence is a weighted average of the true success rate and a "doubt term" (D) that captures the attraction of reversing one's normative preference out of doubt. The latter is a dummy variable taking value 1 if the individual initially believed that she should stop and 0 if she initially believed that she should start further studies.

Let $\bar{q} u(x) + (1 - \bar{q}) u(z) - C(e)$ be the normative EU for the next level. x is the payoff if one succeeds the next level, z is the payoff if one tries the next level and fails, and $C(e)$ is the cost of effort, net of the direct utility of slack time. A rational individual who passed a lower level should stop there if her expected utility of starting the next level is lower than the sure utility of stopping now, conventionally set at 0 value. Having a doubt on this strategy, she becomes overconfident if she should stop ($D = 1$) and underconfident if she should continue ($D = 0$), that is:

$$D = \begin{cases} 1 & \text{if } \bar{q}(\bar{e}, k) u(x) + (1 - \bar{q}(\bar{e}, k)) u(z) - C(\bar{e}) \leq 0 \\ 0 & \text{otherwise} \end{cases}$$

where \bar{e} is the optimal effort when the individual is well-calibrated.

The value taken by the doubt term characterizes an individual's type: either "stopper" if $D = 1$ or "starter" if $D = 0$.

In equation (3.6), the parameter μ ($0 < \mu \leq 1$) indicates the relative precision of the prior (normative) belief. An individual who feels quite sure of her prior preference would resist the temptation of changing her strategy.

Moreover, the same maximization problem can be solved for $\mu = 1$ (well-calibrated) and $\mu < 1$ (miscalibrated). For a given ability level, we obtain a higher value of effort and output for calibrated individuals. We first determine e^* and $D(k, e^*)$ with $\mu = 1$; then, we determine e , q , and E with $\mu < 1$. As a result of miscalibration, some marginal starters will decide to stop and some marginal stoppers will decide to start. The resulting effect on the observed success rate and educational output is ambiguous.

5.3. Performance and effort

In this section, we don't consider the choice of track. All results apply to a given track.

Under the expected utility (EU) criterion, participants define the optimal effort in the task so as to maximize their subjective expected utility net of the cost of effort:

$$\begin{aligned} V &\equiv \max_e q(e, k)u(x) + (1 - q(e, k))u(z) - C(e) \\ &= q(e, k)\Delta u + u(z) - C(e), \text{ with: } \Delta u = u(x) - u(z) > 0 \end{aligned} \quad (3.7)$$

If Δu represents market or experimental exogenous incentives, the maximization of EU for all treatments is equivalent to the maximization of $q(e, k)\Delta u - C(e)$.

With the help of (3.6), we re-write (3.7):

$$V \equiv \mu \rho(ke)\Delta u + (1 - \mu)D\Delta u + u(z) - C(e) \quad (3.8)$$

V is maximized s.t. $e \geq 0$. If the optimal value is $e = 0$, one decides to stop.

We differentiate (3.8) with respect to e to compute the optimal effort for an interior optimum¹⁶. Considering that the cost of effort is a convex function of the efficient time

16. We assume here that the upper bound of effort (8 minutes per round for all rounds) is never

on the task, we set: $C(e) = \frac{1}{2} \times c.e^2$, with: $c > 0$. Since D is exogenously determined, the optimal effort will not depend on the type ("starter" or "stopper").

We summarize¹⁷ below the main effects of the calibration parameter, incentives, cost of effort parameter and ability on effort (3.9) and on the educational output (3.10):

$$e = e\left(\frac{\mu}{c} \Delta u, k\right)_{+} \quad (3.9)$$

$$E = k.e = E\left(\frac{\mu}{c} \Delta u, k\right)_{+} \quad (3.10)$$

Effort and educational output increases with calibration parameter and incentives and decreases with the cost of effort parameter. Although ability has an ambiguous effect on effort, it has always a positive effect on the educational output. We test the implications of the model on our experimental data in Section 7.

Once effort is determined for each track and post-compulsory level, the confidence terms are defined and the student chooses the track (if choice is allowed) that maximizes her EU. The latter decision is examined below.

5.4. The choice of track

Participants in the Choice condition face a decision of track, if succeeding the compulsory level and deciding to continue to post-compulsory levels. The decision of track is based on the comparison of the expected utility for each track. In the VOC, participants must compute their EU for the secondary and tertiary levels, while in the GEN they look forward to the tertiary level since the difficulty, and then chance of

met.

17. The full development of the model is presented in Section A.3 in the Appendix to this Chapter.

success, is the same as in the secondary level. The EU (net of the cost of effort) for each track is given by:

Select VOC and stop after *level 2*:

$$EU_2^V = (1 - q_2^V) \cdot u(y_{20}) + q_2^V \cdot u(y_2) - C(e_2^V) \quad (3.11)$$

Select VOC and go to *level 3*:

$$EU_3^V = (1 - q_2^V) \cdot u(y_{20}) + q_2^V \cdot \left[q_{(3|2)}^V \cdot u(y_3) + (1 - q_{(3|2)}^V) \cdot u(y_{30}) \right] - C(e_2^V, e_3^V) \quad (3.12)$$

Select GEN and go to *level 3*¹⁸:

$$EU_3^G = (1 - q_2^G) \cdot u(y_{20}) + q_2^G \cdot u(y_3) - C(e_2^G, e_3^G) \quad (3.13)$$

where: y_{20} is the outcome if the individual fails secondary level, y_2 is the outcome if she succeeds secondary level and stop, y_{30} is the outcome if she fails tertiary level and y_3 is the outcome for success at tertiary level. For simplicity, we neglect the cost of effort in the following discussion.

Participants select GEN instead of VOC_2 (for the secondary level) if $EU_3^G > EU_2^V$:

$$q_2^G > \left[\frac{u(y_2) - u(y_{20})}{u(y_3) - u(y_{20})} \right] q_2^V \quad (3.14)$$

and GEN instead of VOC_3 (for the tertiary level) if $EU_3^G > EU_3^V$:

18. If decide to GEN, we assume the outcomes for the third level, as the chances of success are the same for second and third levels in this condition (same task and challenge for both levels: $q_{(3|2)}^G = 1$). In other words we can say that in the GEN (rationally) one is able to complete level 3 if he or she is also able to clear the second level.

$$q_2^G > \left[\frac{[q_{(3|2)}^V \cdot u(y_3) + (1 - q_{(3|2)}^V) \cdot u(y_{30})] - u(y_{20})}{u(y_3) - u(y_{20})} \right] q_2^V \quad (3.15)$$

Combining equations (3.14) and (3.15), one selects GEN if:

$$q_2^G > \left[\frac{\max[EU(VOC_{3|2}), u(y_2)] - u(y_{20})}{u(y_3) - u(y_{20})} \right] q_2^V \quad (3.16)$$

6. The effect of financial incentives

Section 5 shows the importance of financial incentives over the behavior and outcomes observed in the task. Decision to start higher levels of the task (equation (3.6)) and selection of the track (section 5.4) depend on the payoff for the tertiary level. Higher are payoffs for upper levels in the task, higher is the likelihood to select the GEN track and to start levels 2 and 3 (if succeeding levels 1 and 2 respectively). These decisions are the main source of the inefficiency observed in the task: (i) the rate of participants starting the secondary level is much higher than the normative chances of success at this level resulting in an excessive level of failure at this level; (ii) the observed inefficiency in the Choice for the secondary level is due to the high, and even excessive, ratio of participants selecting GEN.

Decreasing the payoff for the tertiary level by around 30% has an impact over the expected utility for the whole experiment, influencing not only the decision to perform the upper level of the experiment but also the decision to start the secondary level. Considering participants succeeding the previous level, we have 16% less participants starting the secondary level (61% vs. 77%, $p = 0.0025$) and 32% less participants starting the tertiary level (50% vs. 82%, $p = 0.0025$) in the Choice-low treatment (Table 3.7). Consequently the ratio of failure at secondary level is much lower in the

Choice-low condition (33% vs 50% ¹⁹), representing a "gain" of 18% in the total welfare of Choice-low treatment ²⁰.

Table 3.7 – Incentives and efficiency of school systems

	Choice	Choice-low	<i>p-value</i>
Payoffs (€)	7.80 (7.09)	7.12 (4.48)	0.6949
Output (E)			
Total anagrams solved	54.32 (20.93)	49.17 (17.01)	0.0089
Anagrams solved at compulsory level	41.03 (9.27)	40.22 (7.75)	0.1740
Maximum level attained (l)			
Dropouts (compulsory level)	0.13	0.17	0.3688
Compulsory level	0.70	0.68	0.6161
Post-compulsory levels	0.17	0.15	0.0492
Decision to start next level conditional on success at previous level:			
Secondary level	0.77	0.61	0.0025
Tertiary level	0.82	0.50	0.0093

Notes: **Rate of easy (hard) anagrams solved** considers the 18 anagrams with higher (lower) rate of success in the No-choice treatment among the 36 first anagrams in level 1. **Task specific time** in minutes. **Standard deviation** in parentheses. **Statistical tests:** Payoffs and Task specific time: two-sample Wilcoxon rank-sum (Mann-Whitney) test. Total anagrams solved and Rate of easy (hard) anagrams solved: two-sample t-test with equal variances. Level attained and decisions: two-sample test of proportions.

The self-selection of track is also more efficient in the Choice-low. As we have more participants choosing VOC in the Choice-low (Figure 3.1), the welfare loss caused by the self-selection of the track (section 3.3) is half of that observed with high incentives (12% vs 24% ²¹).

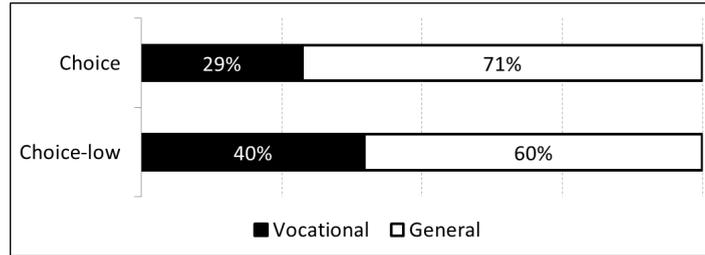
On the other hand, the impact of lower incentives is harmful for the production of anagrams. With lower financial incentives participants solve on average 9.5% less anagrams ($p = 0.0089$) in the entire task (table 3.7).

19. Considering all sample that started the experiment.

20. The difference on payoffs between success at primary level and fail at secondary level is 6 €. If we had in the Choice-low the same rate of failure at secondary observed for Choice (50%), the total welfare in the Choice-low would be 980 € instead of the observed welfare (1,154 €). To compute this value we must assume that individuals that stopped after primary level would not succeed at secondary level. When analyzing the Choice treatment, we can assert that higher incomes to upper levels generated a welfare loss of 12.5% to this treatment, if we had the same ratio of Choice-low participants stopping after primary level, and assuming that those stopping should fail the secondary level, the total welfare in this treatment should jump from 1,584 € to 1,782 €.

21. We have 14 participants selecting GEN and producing enough anagrams to solve the VOC at secondary level, but not the GEN. The difference in payoffs for success and failure at secondary level is 10 €, and the total welfare on the Choice low condition is 1,154 €. In the Choice we have 38 participants selecting GEN that could succeed only in the VOC, and the total welfare in this condition is 1,584 €.

Figure 3.4 – Distribution of participants between the two possible tracks.



Note: participants that started the middle level.

Finally, comparing payoffs for the two choice treatments seems unfair at first sight, but we observe on Table 3.7 that the better efficiency on decisions compensates for the lower payoff at the upper level since the average payoff in the Choice-low is lower but not statistically significant ($p = 0.6949$) than in the Choice.

7. Testing theoretical predictions

7.1. Global performance

In the compulsory level, participants could meet the requirement of solving 36 anagrams over 9 rounds by selecting the easiest anagrams in each round and avoiding thus unproductive time and effort. Besides, all time should be productive in the Screening condition and no time should be wasted in the Race condition. Hence, slack time is minimal at primary level in general ($t \equiv e$). Moreover, in the upper levels, the difficulty is such that subjects have practically no slack time ($t = e$). Overall, time spent on anagrams appears to be a reasonable proxy for effort.

As explained in Section 3, the productivity (speed) of solving anagrams (\bar{k}) is our estimation for ability. Another variable that may explain educational output is the cost of effort (c), for which we do not have a precise measure, but may be positively related to age.

We confirm our theoretical predictions on Table 3.8 presenting an OLS estimation

for the global results in the task, aggregating values for the three different levels. As expected (equation (3.10)), ability ($k \equiv \bar{k}$) induces a higher level of educational output (E). Moreover, the calibration parameter (μ) is positively correlated to the educational output, and the higher is the payoff the higher is the output (Δu , here captured by the Choice-low condition).

Ability (\bar{k}) is negatively correlated to effort (t) which is not inconsistent with the model. Other variables (μ , Δu), as expected, go in the same direction as the educational output.

If on one hand decreasing incentives has a positive effect on the efficiency of decisions (as observed in section 6), on the other hand - as predicted by equations (3.9) and (3.10) - it decreases output (E) and effort (t). In fact, next sections (7.2 and 7.3) show that Choice-low participants don't have lower output (E) and effort (t) at each given level of the task if considering only participants that performed the level (Tables 3.9 and 3.10), but the higher proportion of participants deciding not to start the post-compulsory levels has a direct impact on the global (educational) output (E).

7.2. Compulsory level

Minimizing t (as predicted on section 4.2) is unproductive in the context of this experiment, which explains the inefficiency of Race at the compulsory level. Table 3.9 shows that Race participants spent less time than others at compulsory level, which is the main explanation for the excessive rate of failure in the Race at compulsory level: in order to be fast and "win" the Race, participants forget the main objective of the task, don't devote sufficient effort to solve the maximum of anagrams at each round - at least before reaching the threshold of 36 anagrams to succeed the level -, and fail.

The difference on incentives (Δu) has no effect at compulsory level, since the difference on payoffs is placed at tertiary level. Other variables (μ , \bar{k}) have the same relation to E and t observed in the global performance, those predicted by the theoretical model.

Table 3.8 – Global performance

	Solved anagrams (E)	Task-specific time (t)
Model		
Ability (\bar{k})	45.2292*** (2.2219)	-14.7482*** (1.8572)
Choice-low ($-\Delta u$)	-5.0448*** (1.6699)	-5.4720*** (1.3958)
Precision of reported confidence ($1 - \mu$)	16.4184*** (2.4851)	10.4721*** (2.0771)
Treatments (ref: Choice)		
No-choice	-.3722 (1.5656)	-1.0578 (1.3086)
Screening	1.9331 (1.6021)	1.1820 (1.3391)
Race	-9.1043*** (1.6362)	-8.5999*** (1.3676)
Age	-.2580*** (.0848)	-.2850*** (.0709)
Female	-1.3932 (1.0422)	-1.5701* (.8711)
Risk Averse	-.8242 (1.0442)	-2.2693*** (.8728)
Student	-1.4526 (1.5010)	-.7625 (1.2546)
Constant	38.5101*** (3.5136)	54.0657*** (2.9368)
R^2	.3510	.1861
Observations	941	941

Notes: **Model**: variables correspondent to equations 3.9 and 3.10 presented in Section 5.3. Our proxy for $(1 - \mu)$ is computed by the difference between the reported confidence for the third level before starting the task and the adjusted success rate (adjusted to ability and treatment). **Standard errors** in parentheses. **Significance level**: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

7.3. Post-compulsory levels

When looking at the performance at post-secondary levels on Table 3.10 we observe that, as for the compulsory level, ability (\bar{k}) predicts more anagrams (E) with less effort (t). The precision of reported confidence (μ) also follows the theoretical predictions at post-compulsory levels, but its effect is weaker (and not significant) if compared to the compulsory level or to the global performance in the task.

Individuals with lower incentives spent 13% less time at the secondary level (16.34 vs 18.84 minutes, $p = 0.0089$). The effort (t) for a given level is systematically lower in the Choice-low, but when controlling for the track and other variables, statistically significant only for the secondary level (Table 3.10).

Moreover, we observe that age has a different impact at compulsory and post-

Table 3.9 – Performance at the compulsory level

	Solved anagrams (E)	Task-specific time (t)
Model		
Ability (\bar{k})	13.5850*** (1.0054)	-35.5438*** (1.1561)
Choice-low ($-\Delta u$)	-1.0273 (.7556)	-.4065 (.8689)
Precision of reported confidence ($1 - \mu$)	4.7447*** (1.1245)	.3875 (1.2931)
Treatments (ref: Choice)		
No-choice	-.5894 (.7084)	-.4292 (.8146)
Screening	1.6164** (.7250)	1.0273 (.8336)
Race	-4.7152*** (.7404)	-3.6053*** (.8514)
Age	-.2348*** (.0384)	-.1225*** (.0441)
Female	-.0933 (.4716)	-.6643 (.5423)
Risk Averse	.3822 (.4725)	-.0953 (.5433)
Student	-1.7934*** (.6792)	-.7160 (.7811)
Constant	40.7948*** (1.5899)	45.3342*** (1.8283)
R^2	.2492	.5484
Observations	941	941

Notes: **Model**: variables correspondent to equations 3.9 and 3.10 presented in Section 5.3. Our proxy for $(1 - \mu)$ is computed by the difference between the reported confidence for the third level before starting the task and the adjusted success rate (adjusted to ability and treatment). **Standard errors** in parentheses. **Significance level**: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

compulsory levels since it predicts positively the number of solved anagrams (E) at higher levels of the task. This result suggests that age is a good proxy for the cost of effort (c) for the compulsory level when the level of difficulty of the task is low. But for post-compulsory levels, when difficulty increases, the larger experience/exposure to vocabulary of older people becomes more important to the production of anagrams than the cost of effort for them.

Finally, even if there is no difference on conditions and chances of success among treatments at post-compulsory levels, there is still an impact of them over participants' behavior and performance (Table 3.10). The main impact is observed in the Race, participants replicate the mistaken strategy learned at compulsory level and reinforced by their success at this level, minimizing t even when there is no more competition.

Table 3.10 – Performance at post-compulsory levels

	Secondary level		Tertiary level	
	(E)	(t)	(E)	(t)
Model				
Ability (\bar{k})	7.9727*** (.7856)	-5.1583*** (.8216)	6.9305*** (1.6978)	-8.7569*** (1.8131)
Choice-low ($-\Delta u$)	-.6486 (.6063)	-1.4914** (.6341)	-1.0230 (1.5407)	-.8064 (1.6454)
Precision of reported confidence ($1 - \mu$)	1.1955 (.8800)	-.7979 (.9204)	.6369 (1.9125)	.8641 (2.0424)
Treatments (ref: Choice)				
No-choice	-.8955* (.5313)	-1.6018*** (.5558)	-.7725 (1.1233)	-1.4810 (1.1996)
Screening	.1819 (.5442)	-.4649 (.5692)	-1.5149 (1.2113)	-.5737 (1.2936)
Race	-1.3680** (.5764)	-1.7596*** (.6028)	-.6409 (1.3517)	-1.8427 (1.4436)
Track: GEN (ref:VOC)				
Age	1.8444*** (.3714)	1.6232*** (.3885)	-2.6128*** (.8278)	-1.1520 (.8840)
Female	.0708* (.0364)	-.0823** (.0381)	.1606** (.0760)	.0483 (.0812)
Risk Averse	-.5274 (.3611)	-.3013 (.3777)	-.8265 (.8426)	1.8588** (.8999)
Student	.3060 (.3607)	.0238 (.3773)	-.0749 (.7903)	-2.1517** (.8440)
Constant	.7162 (.5445)	-.1329 (.5695)	.8001 (1.1065)	1.5166 (1.1817)
R^2	7.4192*** (1.3427)	22.7089*** (1.4044)	15.3061*** (2.8910)	22.4883*** (3.0874)
Observations	.2117 595	.1283 595	.1904 134	.2853 134

Notes: **Model:** variables correspondent to equations 3.9 and 3.10 presented in Section 5.3. Our proxy for $(1 - \mu)$ is computed by the difference between the reported confidence for the third level before starting the task and the adjusted success rate (adjusted to ability and treatment). **Standard errors** in parentheses. **Significance level:** * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

7.4. Self-selection of track

The remaining question is *why do participants select GEN when they can opt for the track?* Table 3.11 brings two possible explanations for this question. First, when comparing the subjective chances of success displayed in this table to the normative chances of success in Table 3.3, one can observe that participants are on average overconfident.

Second, same table shows that participants have *limited discrimination*²² about the difference in difficulty among the two tracks. Those performing GEN report on average the same confidence as participants selecting VOC ($q_2^G = q_2^V$). It helps understand

22. See Chapter 1 (subsection 3.1 for more information about limited discrimination when estimating confidence).

why participants select GEN: under the assumption that the cost of effort does not impact the track decision, the conditions to select GEN (equation (3.16)) are met since $q_3^V < q_2^G$. Therefore, few participants chose VOC and those who chose it might have opted for VOC because it implied a lower cost of effort at secondary level. In this case, they had the intention of stopping after passing the secondary level although some of them changed their mind and continued to the tertiary level as their success increased their confidence.

Table 3.11 – Self-confidence for secondary level

	VOC	GEN	<i>p-value</i>
Subjective chances of success	57%	60%	<i>ns</i>

Notes: **Choice** treatment. **Confidence** reported after primary level.
Statistical test: two-sample Wilcoxon rank-sum (Mann-Whitney) test.

8. Conclusion

Can we improve the efficiency of school systems? Is it possible to design school systems that minimize regret? We have proposed an experimental setting with five treatments investigating four different mechanisms to assign students to educational curricula according to their capacity to produce at school: no choice of curriculum, self selection of track, early competition and screening.

When looking at the efficiency of the system, we observe an impact of different mechanisms over the performance of our experimental participants. We find that No-choice and Screening are the more efficient mechanisms, providing higher payoffs, (educational) outcomes and a higher rate of success at tertiary level. Screening results in the highest output (number of solved anagrams) for the primary level as it stimulates sustained effort of individuals at this level.

Early competition (Race) is the worst treatment because participants care not only about their own performance but also about others' performance. In order to *win* the

Race and decide for the track in post-compulsory levels, participants maximize the speed of resolution and forget the central goal of the task, that is, maximize their expected utility. In spite of their high speed, participants in the Race condition presented on average the lowest outcome (E), the lowest payoff, the highest level of failure in the primary level, and the lowest rate of success at tertiary level

The problem of self-selection (Choice) is that it promotes the highest level of failure at secondary level when the economic returns to school are high. We observe that participants selected massively the more difficult track (GEN) and failed. This decision may be ex-ante efficient, because participants maximize their EU, but ex-post inefficient because it brings decreases by 24% the wealth generated in the Choice treatment.

Moreover, higher returns to tertiary education increase the amount of regret. Choice-low participants had a lower rate of individuals deciding to start the secondary level, and consequently a lower rate of failure at this level. Moreover, a lower expected utility for the experiment resulted in a more efficient tracking at post-compulsory levels, the observed welfare loss due to the self-selection of track in the Choice-low is half of that observed in the Choice. If on one hand lower incentives result in more ex-post efficient decisions, on the other hand it induces participants to lower levels of effort and consequently lower output (number of solved anagrams).

These findings may explain the high level of dropout and failures on higher levels of education. The inefficiency of the system derives from two main reasons: (i) if students have an imperfect knowledge of their own ability, and/or if there is a lack of discrimination between the two tracks (Vocational and General), they are inclined to opt for the more difficult track and fail; (ii) the higher is the wage premium for tertiary education the higher is the ex-ante expected utility, what boosts the chances to try higher levels of education, but does not increase the probability of success, raising the ex-post inefficiency (higher level of failures, dropouts and regret).

Several clear policy implications emerge from our results in the challenge to design

a system that is ex-post efficient. First, screening students by ability seems to be the best mechanism to track students by ability, the question is to set grades (thresholds) that encourage less able but motivated students to reach higher levels of education, without discouraging the less motivated to complete the primary level.

Second, we believe that the efficiency of the No-choice is directly linked to statistical issues. We do not advocate that this is the best solution for educational institutes, as we believe that students must choose their preferred track, but we cannot neglect that educational institutions must also support students in the orientation choices, since self-selection has an important weight in the educational decisions even if institutes filter students by ability level. Making educational choices is hard, since when deciding the student does not know (perfectly) her ability and the difficulty that she will face in the future. The single decision of preferred track brought a welfare loss of 12% in our experiment. Policymakers must design not only efficient sorting mechanisms, but help students understand the difficulties they will face in the future, reducing the dropout and failures on higher levels of education. It seems crucial to evaluate on the gateway not only performance but also intrinsic motivation and other non-cognitive abilities.

Third, early competition must be avoided. Under competition students worry about others' performance leaving aside the own educational output, which should be the major goal of education.

External validity and limitations. Our experimental design has two main limitations when compared to a real school system: *(i)* we don't observe differences in opportunities, *(ii)* impatience and discount rates are not observed because schooling time is compressed in our experiment. However, we don't have the ambition of reproducing the school system with all its characteristics and complexities in the laboratory. First because we aim at identifying a specific causal relationship, second because reproducing perfectly a school system is impossible, and finally even in the case where it is possible, because we would use real data instead since they are more natural. According

to Camerer (2015), reproducing all characteristics of the environment studied is not a primary concern in a typical experiment, since experimental economics aims at establishing a general theory linking economic factors such as incentives, rules and norms to decisions and behavior. However, our findings are coherent to literature, suggesting that our experimental framework is externally valid. For instance, in a randomized controlled trial in France, Goux, Gurgand and Maurin (2016) show that increasing the quality of information about students' ability and describing better the difference of the two possible tracks²³ improves the quality of school tracking, and reduces failures at high school by 25% in the target population. They explain the same observation we found in the Choice treatment (explaining the choice paradox) with a different methodology. Another example comes from Filippin and Paccagnella (2012) when comparing school tracking in Netherlands' and Italy's school systems. The main difference between this two countries concerns the self-selection for the high school tracks²⁴. In Netherlands, results of a nationwide aptitude test at age 12 act as a reference of the most suitable track for pupils aptitudes. Differently, pupils -and their parents- have no signal to select their preferred track in Italy. This comparison show the effectiveness in the quality of tracking students by ability in Netherlands. We find the same pattern of result when comparing Screening (that is closer to Netherlands' system) and Choice (that is closer to Italy's system).

23. In France, students must decide for the high school track (vocational or academic) by the end of middle school

24. The three possible tracks here are academic (general), vocational and technical.

A. Appendices

A.1. Tables and figures

Table 3.12 – Efficiency of treatments (complete version)

	No-choice	Choice	Screening	Race	Choice-low
Payoffs (€)	9.00	7.80	8.18	7.22	7.12
Output (E)					
Total anagrams solved	54.65	54.32	56.38	48.42	49.17
Anagrams solved at compulsory level	40.59	41.03	42.64	37.22	40.22
Anagrams solved at secondary level	9.54	10.19	10.16	8.67	7.36
Anagrams solved at tertiary level	4.51	3.09	3.58	2.52	1.58
Maximum level attained (l)					
Fail compulsory level	0.13	0.13	0.14	0.23	0.17
Compulsory level	0.63	0.70	0.64	0.59	0.68
Secondary level	0.10	0.06	0.11	0.10	0.10
Tertiary level	0.14	0.11	0.11	0.08	0.05
Decision to start next level conditional on success at previous level:					
Secondary level	0.77	0.77	0.79	0.79	0.61
Tertiary level	0.82	0.82	0.76	0.63	0.50

Table 3.13 – Statistical tests for Table 3.12

p – value	No-Choice			Choice		Screening	Choice-low
	Choice	Screening	Race	Screening	Race	Race	Choice
Payoffs (€)	0.0948	0.3428	0.0049	0.5032	0.1633	0.0572	0.6949
Output (E)							
Total anagrams solved	0.7672	0.1449	0.0032	0.2090	0.0016	0.0000	0.0089
Anagrams solved at compulsory level	0.2271	0.0007	0.0000	0.0671	0.0000	0.0000	0.1740
Anagrams solved at secondary level	0.4127	0.4262	0.2815	0.9056	0.0609	0.0742	0.0014
Anagrams solved at tertiary level	0.0864	0.3415	0.0161	0.4462	0.4462	0.1329	0.0492
Maximum level attained (l)							
Fail compulsory level	0.8235	0.6296	0.0075	0.7937	0.0145	0.0316	0.3688
Compulsory level	0.0491	0.5523	0.1074	0.1796	0.7713	0.3113	0.6161
Secondary level	0.2667	0.3068	0.0623	0.9456	0.4183	0.3878	0.0418
Tertiary level	0.2667	0.3068	0.0623	0.9456	0.4183	0.3878	0.0418
Decision to start next level conditional on success at previous level:							
Secondary level	0.9148	0.7008	0.6418	0.7812	0.7164	0.9231	0.0025
Tertiary level	1.0000	0.4657	0.0441	0.5151	0.0725	0.2049	0.0093

Statistical tests: Payoffs: two-sample Wilcoxon rank-sum (Mann-Whitney) test. Total anagrams solved: two-sample t-test with equal variances. Level attained and decisions: two-sample test of proportions.

Figure 3.5 – Maximum level attained (cumulative distribution)

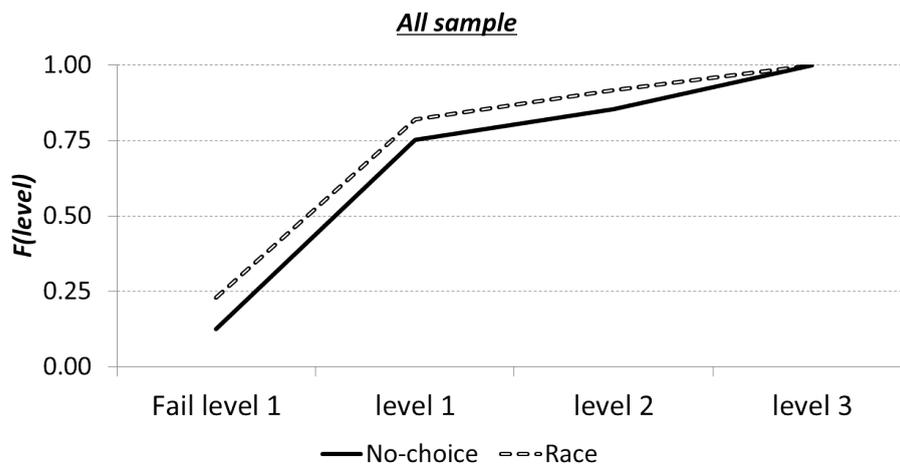
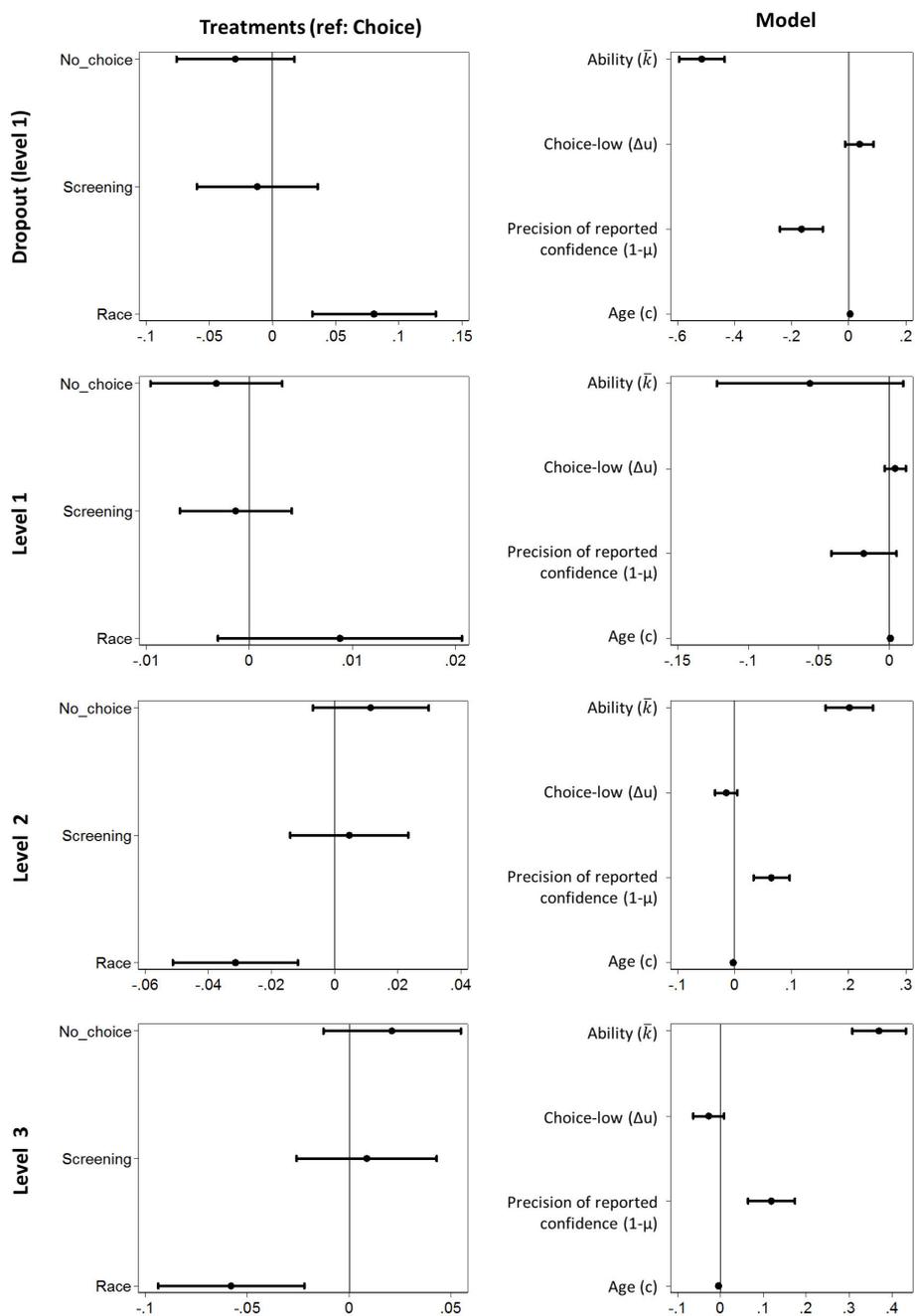


Table 3.14 – Maximum level attained: probability of success

Ordered probit	Probability to attain levels (0, 1, 2, 3)
<i>Model</i>	
Ability (\bar{k})	2.5478*** (.1829)
Choice-low ($-\Delta u$)	-.1878 (.1260)
Precision of reported confidence ($1 - \mu$)	.8196*** (.1883)
<i>Treatments</i> (ref: Choice)	
No-choice	.1453 (.1180)
Screening	.0592 (.1210)
Race	-.3980*** (.1236)
Age	-.0291*** (.0067)
Female	-.2155*** (.0785)
Risk Averse	.0649 (.0783)
Student	-.1487 (.1136)
Cut 1	-.7973 (.2700)
Cut 2	1.4510 (.2727)
Cut 3	1.9521 (.2779)
Pseudo R^2	0.1430
Observations	941

Notes: **Model**: variables correspondent to equations 3.9 and 3.10 presented in Section 5.3. Our proxy for $(1 - \mu)$ is computed by the difference between the reported confidence for the third level before starting the task and the adjusted success rate (adjusted to ability and treatment). **Level 0**: dropout in the compulsory level **Standard errors** in parentheses. **Significance level**: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 3.6 – Maximum level attained: marginal effects (table 3.14)



A.2. Decisions after level 1

Individuals who succeeded the first level have 4 possible outcomes when considering the Choice treatment. A rational agent is supposed to select the option that maximizes the expected utility (net of the cost of effort if they continue) comparing the 4 possibilities. According to section 5.4, the decision to start the secondary level followed by the track decision is given by the comparison of the following expected utilities:

- [1] Stop after level 1:

$$EU_1 = u(y_1)$$

- [2] Select VOC and stop after level 2:

$$EU_2^V = (1 - q_2^V) \cdot u(y_{20}) + q_2^V \cdot u(y_2)$$

- [3] Select VOC and go to level 3:

$$EU_3^V = (1 - q_2^V) \cdot u(y_{20}) + q_2^V \cdot \left[q_{(3|2)}^V \cdot u(y_3) + (1 - q_{(3|2)}^V) \cdot u(y_{30}) \right]$$

- [4] Select GEN and go to level 3²⁵:

$$EU_3^G = (1 - q_2^G) \cdot u(y_{20}) + q_3^G \cdot u(y_3)$$

Table 3.15 present the chances of success for GEN and VOC in the No-choice treatment. We consider the chances of success in the No-choice to avoid the effect of other treatments in the chances of success for higher levels, anyway results are not qualitatively different when considering the chances of success for all treatments.

Table 3.15 – Ex-post chances of success for No-choice participants

Ability level:	track	
	VOC	GEN
q_2	43%	31%
q_3	19%	24%

Table 3.16 present the EU for each possible decision. The conclusion is that, given the ex-post chances of success, participants should select the general track and try to

25. Assuming outcomes for the third level, as the chances of success are the same for second and third levels in the GEN (same challenge for both levels) $q_{(3|2)}^G = 1$.

reach the tertiary level, with the exception of risk averse individuals whose should stop after level 1 ²⁶.

Table 3.16 – Expected utility for different ability levels and risk profiles

	Risk neutral	Risk seeker (y^2)	Risk averse (\sqrt{y})
EU_1	10	100	3.16
EU_2^Y	8.30	93.40	2.75
EU_3^Y	9.86	166.60	2.90
EU_3^G	10.82	220.60	2.96

26. The decision to stop after first level, if participant clear the primary level in the Choice treatment does not differ according to our measure of risk aversion. The proportion of risk averse participants stopping after primary level is higher than for risk seekers (25% vs 21%), but not statistically significant at 10% level ($p = 0.5236$: two-sample test of proportions).

A.3. Performance and effort (complete version)

In this section, we don't consider the choice of track. All results apply to a given track.

Under the expected utility (EU) criterion, participants define the optimal effort in the task so as to maximize their subjective expected utility net of the cost of effort:

$$\begin{aligned} V &\equiv \max_e q(e, k)u(x) + (1 - q(e, k))u(z) - C(e) \\ &= q(e, k)\Delta u + u(z) - C(e), \text{ with: } \Delta u = u(x) - u(z) > 0 \end{aligned} \quad (3.17)$$

If Δu represents market or experimental exogenous incentives, the maximization of EU for all treatments is equivalent to the maximization of $q(e, k)\Delta u - C(e)$.

With the help of (3.6)²⁷, we re-write (3.17):

$$V \equiv \mu \rho(ke)\Delta u + (1 - \mu)D\Delta u + u(z) - C(e) \quad (3.18)$$

V is maximized s.t. $e \geq 0$. If the optimal value is $e = 0$, one decides to stop.

We differentiate (3.18) with respect to e to compute the optimal effort for an interior optimum²⁸. Considering that the cost of effort is a convex function of the efficient time on the task, we set: $C(e) = \frac{1}{2} \times c.e^2$, with: $c > 0$. Since D is exogenously determined, the optimal effort will not depend on the type ("starter" or "stopper"). The first order condition shows that:

27. This equation is presented in page 160.

28. We assume here that the upper bound of effort (8 minutes per round for all rounds) is never met.

$$\begin{aligned}\mu k \rho'(ke) \Delta u &= C'(e) \\ &= c.e\end{aligned}\tag{3.19}$$

$$e = \left[\mu k \rho'(ke) \right] \frac{\Delta u}{c}\tag{3.20}$$

Effort increases with the calibration parameter and incentives and decreases with the cost of effort parameter. To see how effort varies with ability, we differentiate (3.20) with respect to k :

$$\frac{\partial e}{\partial k} = \rho''(ke) \left[e + k \frac{\partial e}{\partial k} \right] \frac{\mu}{c} k \Delta u + \rho'(ke) \frac{\mu}{c} \Delta u$$

Rearranging terms, we get:

$$\begin{aligned}\frac{\partial e}{\partial k} \left[1 - \rho''(ke) \frac{\mu}{c} k^2 \Delta u \right] &= \rho''(ke) \frac{\mu}{c} k \Delta u \times e + \rho'(ke) \frac{\mu}{c} \Delta u \\ &= \frac{\mu}{c} \Delta u \rho'(ke) \left[1 + \frac{\rho''(ke)}{\rho'(ke)} k e \right]\end{aligned}$$

With the help of (3.20), this may be rewritten:

$$\frac{\frac{\partial e}{\partial k}}{\frac{e}{k}} = \frac{1 + \frac{\rho''(E)}{\rho'(E)} E}{1 - \frac{\rho''(E)}{\rho'(E)} E}$$

that is, the elasticity of effort with respect to ability, and yields:

$$\frac{\partial e}{\partial k} \begin{matrix} \leq \\ > \end{matrix} 0 \quad \text{if} \quad \frac{\rho''(E)}{\rho'(E)} E \begin{matrix} \leq \\ > \end{matrix} -1$$

Thus, ability has an ambiguous effect on effort. We summarize below the main effects:

$$e = e\left(\frac{\mu}{c} \Delta u, k\right) \quad (3.21)$$

Although ability has an ambiguous effect on effort, it has always a positive effect on the educational output, as shown below:

$$\frac{\partial E}{\partial k} = \frac{\partial(ke)}{\partial k} = e + k \frac{\partial e}{\partial k} = e + \frac{1 + \frac{\rho''}{\rho'} E}{1 - \frac{\rho''}{\rho'} E} \times e = \frac{2e}{1 - \frac{\rho''}{\rho'} E} > 0$$

$$\frac{\frac{\partial E}{\partial k}}{\frac{E}{k}} = \frac{2}{1 - \frac{\rho''}{\rho'} E} < 1 \quad , \text{ if } \frac{\partial E}{\partial k} < 0$$

and the elasticity of educational output with respect to ability is lower than one if $\frac{\partial E}{\partial k} < 0$ and higher than one if $\frac{\partial E}{\partial k} > 0$.

We summarize below the main effects on educational output:

$$E = k.e = E\left(\frac{\mu}{c} \Delta u, k\right) \quad (3.22)$$

We can then conclude that effort, measured by the time observed in the task, is a function of incentives (increasing) and endowed ability (increasing or decreasing),

whereas the expected (educational) output, measured by the number of solved anagrams, is an increasing function of incentives and endowed ability, as described by equations (3.21) and (3.22). These are testable implications of the model on our experimental data.

CHAPTER 4

**SORTING AND SOCIOECONOMIC BIAS OF
SCHOOL SYSTEMS**

This chapter is a joint work with Louis Lévy-Garboua.

1. Introduction

School systems are socioeconomically biased¹ as children born in the upper classes of society are considerably overrepresented in the upper levels of education. They may also be gender-biased. The socioeconomic and gender bias of school systems is a major concern for the democratic societies of advanced countries whose social contract relies on the consensus that chances should be equal for all. Economists and other social scientists have been describing this phenomenon for several decades (Boudon, 1973, Jencks, 1973, Becker and Tomes, 1979, Bourdieu and Passeron, 1984, Baudelot and Establet, 1992, Breen and Goldthorpe, 2001, Hanushek and Wößmann, 2006, Ruhose and Schwerdt, 2016, among many others) and public policies have been trying to remedy it relentlessly. The classical analysis of Becker (1967) on intergenerational inequality attributes the inequality of chances essentially to differences in abilities and opportunities. Considering that innate abilities are equally distributed in all social classes, the socioeconomic bias of school systems should vanish once social differences in opportunities can be neutralized. Human capital theory (Becker, 1964) demonstrates that an efficient credit market on education investments is all that we need to reach this goal. This optimistic prediction has not quite materialized in developed countries, however, in spite of sustained efforts to eradicate differences in opportunities.

Several studies have shown that differences in opportunities play only a marginal role nowadays in developed countries (Carneiro and Heckman, 2002, Cameron and Taber, 2004). Thus, *why does the socioeconomic bias persist in school systems of developed countries and why does it differ between countries and periods?* This chapter is an attempt to answer this important question in a novel manner. The persistence of the socioeconomic bias of school systems demonstrates the presence of socioeconomic-biased noncognitive abilities, built up during childhood and adolescence. Those noncognitive

1. We define socioeconomic bias in education by the degree to which educational decisions and educational attainment is impacted by pupil's socioeconomic status.

abilities are inherited by children and youth from their permanent exposition to their parents, friends, peers, and social environment and the differential investment of families in their human capital (Lévy-Garboua, 1973, Becker and Tomes, 1979, Cunha and Heckman, 2008, Heckman and Farah, 2009, Cunha, Heckman and Schennach, 2010). Considering that innate abilities are equally distributed in all social classes, noncognitive abilities unequally distributed between social classes generate a socioeconomic bias if school systems perform unequally across ability levels. Our contribution here is to demonstrate the *interaction* of such noncognitive abilities with the sorting mechanisms of current school systems and with the economic returns to education. The origin and nature of such abilities being immaterial, our empirical strategy is to rely on experimental data for measuring the specific ability to perform a task of increasing difficulty, and on *minimal school systems*² to compare the interaction of several sorting mechanisms with such measured ability. This methodology enables us to compare the socioeconomic and gender biases of various school systems.

We compare *four mechanisms* for sorting students according to their abilities: self-selection of further studies with a single track (no-choice of track), self-selection of further studies with the choice of track, screening by ability and early *numerus clausus* competition. We find that competition is the worst institution for high and medium-ability individuals, while self-selection of track is the worst treatment for low-ability individuals. We also compare school choices under two incentive structures: high versus low returns to tertiary education. Our basic assumption for interpreting these results is that students do not know perfectly their own ability prior to making educational choices and thus they are not sure to reach their desired level of education. Consequently, they make decisions that they will often regret in the future. Students who failed and suffered an opportunity loss will regret their choice ex-post and ask for political redistribution. While political platforms aim at reducing failures -viewed as

2. The minimal school system is described in the Introduction to Part II (page 126).

wastage, or ex-post inefficiency-, students pursue their own objective of maximizing expected utility (EU) which, unfortunately, does not guarantee future success and may cause regret. The question studied in this chapter is: *how do different school systems and school returns affect differently ability groups, genders, and social groups, thus causing substantial differences in social and gender bias among developed countries and periods?*

This paper is organized as follows: Section 3 examines the interaction of prior noncognitive ability with sorting mechanisms and how it affects the socioeconomic bias of school systems. Sub-section 3.2 compares the performance of sorting mechanisms for each ability level, whereas sub-section 3.3 compares the relative performance of ability groups for each sorting mechanism. Section 4 numerically illustrates why school systems are socioeconomically biased even if all students have equal opportunities. Section 5 is devoted to the description and discussion of the gender bias of sorting mechanisms. Section 6 shows the surprising socioeconomically biased effects of a rise in the returns to tertiary education. The main conclusions and policy implications of our experimental analysis of school systems are discussed in section 7.

2. Experimental design, experimental sessions and descriptive statistics

The procedure of the experiment and details of the experimental sessions are described in the introduction to Part II (page 127). Descriptive statistics of our 941 participants are also provided in the introduction to Part II (page 133).

3. On the interaction of ability with sorting mechanisms and how it affects the socioeconomic bias of school systems

3.1. Ability level

The main objective of the task is to solve two thirds of the proposed anagrams in a limited amount of time. The (normalized) number of anagrams solved per minute in the first four rounds of the primary level is our estimated ability to perform the task³. We select the measure in the first four rounds because it is supposedly an exogenous⁴ and good measure of prior ability, which captures all noncognitive prior effects of family and the environment on the student's capacity of learning (cognitive ability).

We group participants in three different ability terciles (per treatment): high, medium and low. Some descriptive statistics for our ability groups are reported in Table 4.1. The three groups are homogeneous in terms of gender and risk aversion but a slightly greater proportion of low-ability subjects can be found among older, probably non-student, participants.

Ability strongly discriminates among participants in terms of performance (solved anagrams and payments) and dropouts before the secondary level. However, the primary level was meant to be easy enough that three-quarters (190:263) of low-ability subjects would pass it.

3. We normalize the ability dividing values by 6, that is the highest number of anagrams solved per minute in our sample. As the maximum of anagrams solved per minute is 6, our normalized measure is equivalent to the number of anagrams solved each 10 seconds in the four first rounds of the primary level.

4. This measure is not completely exogenous because participants decide the level of effort devoted to the task. The level of effort devoted to the task may be influenced by treatments, but it is not a problem for our analysis because we compare ability only intra-treatment, for instance when grouping participants by ability level.

Table 4.1 – Descriptive statistics by ability level

Ability level:	High (HA)	Medium (MA)	Low (LA)
Ability	0.76 (0.15)	0.41 (0.08)	0.20 (0.07)
Women	0.47	0.48	0.52
Age	23.57 (5.76)	24.26 (7.22)	26.65 (9.71)
Risk Averse	0.54	0.50	0.58
Student	0.80	0.83	0.77
Payoff	10.90 (9.16)	7.48 (6.59)	5.98 (4.36)
Total anagrams solved	65.34 (19.59)	53.12 (17.70)	42.70 (15.51)
Rate of success at training level	0.97	0.85	0.72
Observations	252	264	263
Decision to start next level conditional on success at previous level:			
Middle level	0.91	0.80	0.58
High level	0.85	0.68	0.56

Notes: No-choice, Choice, Screening and Race treatments. Terciles for ability level are not strictly equal because they were first computed within-treatment to avoid a potential effect of treatment, then aggregated across treatments. Number of participants successfully clearing level 1/level 2: HA: 244/93, MA: 93/50, LA: 190/16. Standard deviation in parentheses. Statistical tests are reported in table 4.2.

Table 4.2 – Descriptive statistics by ability level: statistical tests for Table 4.1

<i>p</i> – value	High-ability		Medium-ability	Test
	Medium-ability	Low-ability	Low-ability	
Ability	0.0000	0.0000	0.0000	(ii)
Women	0.6464	0.2279	0.4510	(iii)
Age	0.1709	0.0000	0.0027	(ii)
Risk Averse	0.7104	0.1860	0.1876	(iii)
Student	0.1849	0.3371	0.1214	(iii)
Payoff	0.0000	0.0000	0.0234	(i)
Total anagrams solved	0.0000	0.0000	0.0000	(ii)
Rate of success at training level	0.0000	0.0000	0.0004	(iii)
Decision to start next level conditional on success at previous level				
Middle level	0.0005	0.0000	0.0274	(iii)
High level	0.0704	0.0076	0.1968	(iii)

Statistical tests: (i) Two-sample Wilcoxon rank-sum (Mann-Whitney) test, (ii) Two-sample t-test with equal variances, (iii) Two-sample test of proportions.

3.2. How do school sorting systems interact with ability?

We have three main criteria to evaluate participants' performance in the task: payoff and output (measured by the number of solved anagrams), and the maximum level attained. The higher are payoffs, outputs and rate of success at post-compulsory levels, and the lower are the dropout rates (failures), the higher is the productive efficiency of the treatment.

Result 1: High-ability students perform equally well under all mechanisms with the notable exception of Race under which they perform relatively badly. Medium-ability and low-ability students are negatively affected by two sorting mechanisms: Race and Choice. However, Choice is particularly harmful to the low-ability group. No-choice and Screening perform relatively well for the three terciles of ability. Early competition is especially harmful in primary education whereas the choice of track causes regret in secondary education. In the Race, high-ability individuals (HA) perform worse than in other treatments. Table 4.3 shows how performance varies with sorting mechanisms for each ability level. In this table, the dropout rate of HA at secondary level is 15% higher in the Race⁵. The consequence of this inefficiency is drastic: in the Race, HA individuals have on average lower payoffs relative to No-Choice and Choice, lower number of solved anagrams among all treatments, and lower ratio of participants achieving the secondary or tertiary levels than in No-choice (Table 4.3).

Race is also inefficient to medium-ability (MA) participants because the dropout rate at compulsory level is at least twice higher than in other treatments. It seems that competition discourages or disturbs the performance of MA individuals pushing them to perform like low-ability participants (LA) at the compulsory level. The problem of Choice for MA is the dropout rate at secondary level. Consequently, the gap on payoffs

5. 70% in the Race vs. 55% on average for other treatments.

Table 4.3 – Variation of performance across sorting mechanisms by ability level

	No-choice	Choice	Screening	Race
High ability (HA)				
Payoff (€)	11.96 (9.55)	11.37 (9.21)	10.93 (9.12)	9.14 (8.64)
Total anagrams solved	66.38 (20.14)	69.19 (21.06)	65.93 (19.18)	59.44 (16.71)
Dropout rate at compulsory level	.03	.02	.05	.03
Dropout rate at secondary level ⁶	.51	.56	.57	.70
Attainment rate at secondary or tertiary levels	.45	.39	.35	.27
Medium ability (MA)				
Payoff (€)	8.45 (7.62)	6.94 (6.04)	7.74 (6.61)	6.73 (5.92)
Total anagrams solved	53.16 (18.60)	54.22 (16.24)	57.97 (17.70)	46.73 (16.86)
Dropout rate at compulsory level	.14	.07	.11	.30
Dropout rate at secondary level	.67	.84	.68	.64
Attainment rate at secondary or tertiary levels	.22	.12	.24	.18
Low ability (LA)				
Payoff (€)	6.58 (4.34)	5.47 (4.13)	6.01 (4.32)	5.83 (4.70)
Total anagrams solved	44.42 (14.04)	40.87 (15.67)	45.91 (16.52)	39.27 (15.17)
Dropout rate at compulsory level	.20	.31	.26	.35
Dropout rate at secondary level	.83	.96	.83	.79
Attainment rate at secondary or tertiary levels	.07	.01	.07	.08

Notes: Standard deviation in parentheses. Statistical tests are reported in the Appendix (Table 4.10, page 212).

for HA and MA is higher in the Choice condition: MA obtain there only 61% of what HA get versus 72% for other treatments.

Payoffs, output and success rate at compulsory level for LA in the Choice and Race treatments are very similar and worse than No-choice and Screening. Choice is especially unfair to low-ability individuals because the latter present a huge dropout at secondary level and the lowest attainment rate at post-secondary levels, only 1% of low-abilities succeed beyond the compulsory level in this condition.

These results show that Race is inefficient for all ability levels. The competition for the right to select one's preferred track in post-compulsory levels interferes on the strategy to perform the task. Participants are not only concerned with their own performance, but also with the performance of other participants in the same experimental session. In order to *win* the Race, participants minimize the time of resolution of anagrams instead of maximizing the number of solved anagrams which is the main objective of the task. Indeed, participants spend on average 20 minutes at compulsory level in the Race, while participants in other treatments use on average 26 minutes at this level ($p = 0.0000$, two sample t-test).

This mechanism disturbs the performance of all participants in the Race, with a contrasted impact for different ability levels: LA and MA fail more than in other treatments at primary level, and HA fail more at secondary level. The inefficiency appears later for HA because the primary level is easy for them. Note however that there is no more competition at level two, so no reason to perform differently than in other treatments. A possible explanation for low performance of HA may then be that they replicate at the secondary level the strategy learned at compulsory level, i.e. solving anagrams faster than others. The ex-post result confirms that Race participants do not adopt a good strategy since outcomes in the task do not rely on relative performances.

3.3. Which school system is more socioeconomically biased?

We compare now the relative performance of ability groups across treatments. Performance is measured successively by payoffs, number of anagrams solved, and distribution of highest level successfully completed. The more favorable to upper ability groups is a sorting mechanism, the more socioeconomically biased it is. The socioeconomic bias of a school system means that disadvantaged groups have lower access to education and to the higher occupations and social positions.

Result 2: In terms of payoffs and educational output, Choice generates a relatively high inequality. Race appears to cause lower socioeconomic bias than other sorting mechanisms. This is actually a direct consequence of the relative inefficiency of this mechanism. Figure 4.1 shows that with regard to payoffs, Choice generates a relatively high inequality. Race appears to cause lower socioeconomic bias than other sorting mechanisms. This is actually a direct consequence of the relative inefficiency of this mechanism (see Section 3.2, Table 4.3). Since Race was especially harmful to HA and MA, it appears to generate less inequality. The equalizing virtue of the Race condition is paid by the opportunity loss incurred by upper groups.

Concerning educational output, Choice only produces inequality. LAs solve on average 40% less anagrams than HAs in the Choice treatment. Screening and Race generate no more inequality than No-choice⁷.

Figure 4.1 – Ratio of average payoffs and outputs (low/high ability)

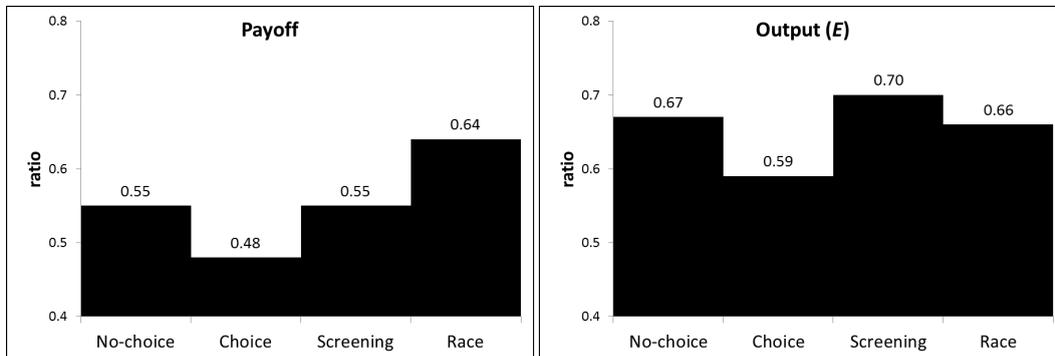
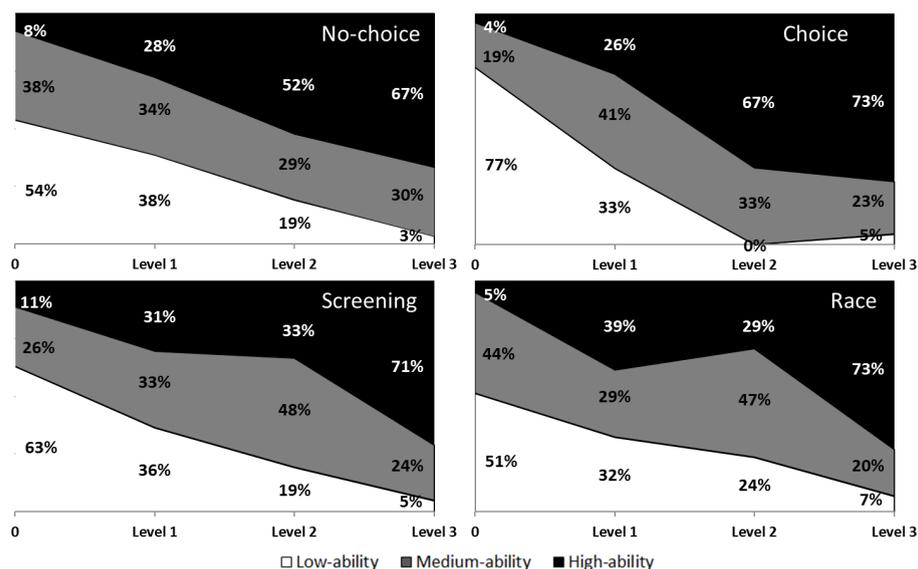


Figure 4.2 visualizes the ability-biasing effects of the four sorting mechanisms considered here. If there was no bias with a sorting mechanism, the three ability terciles would have an equal share at all levels of education, and horizontal parallels would partition the corresponding box in three colored areas of equal surface. Obviously, the four mechanisms manage to sort participants by their ability as can be expected: the dividing lines decline quite abruptly and the share of upper ability groups rise with the level. However, the patterns of such decline are quite different across treatments. Remarkably, the No-choice treatment exhibits almost no ability (socioeconomic) bias for the MA group since the latter's share is about one-third at all levels. In comparison, Choice is harmful and Screening is helpful for this group at secondary level, whereas Race especially hurts the MA in the compulsory level. While the No-choice system is neutral on MA, it smoothly but powerfully sorts HA from LA. In comparison, Choice practically eliminates LA from secondary and tertiary levels.

7. Payoff-ratio: differences between Choice and Race / Screening and Race are significant at 1% level, all other differences are not significant at 10% level levels. Output: difference between Choice and No-Choice, Choice and Screening, and Choice and Race are significant at 5% level, all other differences are not significant at 10% level. (Wilcoxon signed-ranks).

Figure 4.2 – Participation of ability groups at each level attained in the task



Note: Level 1: succeeds the first level, does not start or fails the second level. Level 2: succeeds the second level, does not start or fails the third level.

Screening and Race both seek to sort students by letting the institution observe a specific ability index. However, their different conception of "ability" is reflected by the difference in their rewarding policies. Screening stimulates personal motivation by rewarding cumulated output and the sustainment of effort whereas Race stimulates rivalry by rewarding speed and the intensity of a short effort. Consequently, Screening permanently benefits the motivated students notwithstanding their ability group whereas Race initially benefits the faster runners of the first tercile but gives them a bad signal in the long run. An unexpectedly happy consequence of the harmful effect of early competition on able students is that Race, while being inefficient, tempers the socioeconomic-biasing effect of school.

4. The socioeconomic-biasing effect of choice explained

The socioeconomic-biasing effect of Choice is a major result elicited in the previous section. Choices of further education and track interact with ability in ways that amplify the strength of students' sorting by ability, thus exacerbating prior noncognitive ability differences by social origin.

Although economic thinking is prone to highlight the benefits of choice, some behavioral distortion may push students into making bad decisions. To see this, let us compare what participants did in our experiment with what a rational EU-maximizer should do.

Notice first that general studies should be pursued up to the tertiary level since, if you passed the secondary level, you should pass the tertiary level too which is no more difficult. Therefore, the list of potential schooling decisions for a rational student who passed the primary level is the following: stop after primary education, take secondary vocational studies and stop, take vocational studies up to the tertiary level, take general studies up to the tertiary level. Table 4.4 presents the expected utility⁸ of all these potential decisions to rational and well-calibrated students who would know the exact probabilities of transition for all relevant tracks and levels given their own ability group.

The decision may depend on the individual's risk attitude and cost of effort. In Table 4.4, we neglect the cost of effort and focus on risk attitude. Three values of constant relative risk aversion are considered: 0 (risk neutrality), -1 (risk-seeking), and 0.5 (moderate risk aversion). What Table 4.4 shows, then, is that, in our experiment, high-ability subjects should always try to reach the tertiary level, whereas low-ability subjects should always stop after the compulsory level. Medium-ability subjects would have a more heterogeneous behavior, with the risk-tolerant behaving like HA and the risk-averse behaving like LA. Besides, it never pays to stop after the secondary level.

8. For more information, the model of rational decisions is presented in the Appendix to this chapter, Section A.2 (page 215).

Table 4.4 – Expected utility for different ability levels and risk profiles

Expected utility	Risk neutral	Risk seeking (y^2)	Risk averse (\sqrt{y})
High-ability			
Stop after level 1	10	100	3.16
Level 2 in the VOC	9.70	118.60	2.99
Level 3 in the VOC	12.49	242.35	3.29
Level 3 in the GEN	13.24	293.20	3.30
Medium-ability			
Stop after level 1	10	100	3.16
Level 2 in the VOC	7.60	80.80	2.63
Level 3 in the VOC	8.92	142.60	2.76
Level 3 in the GEN	10.38	207.40	2.90
Low-ability			
Stop after level 1	10	100	3.16
Level 2 in the VOC	6.30	57.40	2.40
Level 3 in the VOC	5.61	40.15	2.30
Level 3 in the GEN	6.64	95.20	2.37

Note that it is always better to seek tertiary education through the general track, with the exception of risk-averse high-ability subjects who are almost indifferent between the two tracks. This result is consistent with the fact that screening especially motivates medium and low ability individuals relative to high-ability ones, thus exerting an equalizing effect.

The predictions that derive from Table 4.4 must now be compared with the real decisions. Table 4.5 presents the observed frequencies of quitting after the primary level and track selection for the three ability groups as a function of the school system. As decisions may depend on the risk attitude, the EUs were computed for a reasonable range of CRRA parameters. First, Table 4.5 shows that the decision to stop after compulsory level declines sharply when ability rises, with a significant difference between risk attitudes for a given ability group. Since real decisions are based on subjective estimates of future success, we conclude that the 58.7% of LA and 73.0% of risk averse-MA who started the secondary level while it was rational for them to stop were overconfident. Likewise, the 8.8% of HA who stopped while it was rational for them to continue were underconfident. These two conclusions illustrate on our data the Dunning-Kruger effect (Kruger, 1999) studied in Chapter 1. However, LAs pursuing education too far are in much greater proportion than HAs stopping education too

early; and the excessive volume of failures resulting from this situation may be a first cause of aggravation for the socioeconomic bias initiated by the unequal distribution of prior noncognitive abilities. Second, we observe in Table 4.5 that all risk-seeking and the risk-averse MAs behave differently when selecting the track for post-compulsory level: a smaller proportion of risk-averse MAs, supposed to stop after level 1 (Table 4.4) choose GEN for the post secondary levels. Differences in the track selection for different risk attitudes are not statistically different at usual levels for HAs and LAs.

Table 4.5 – Decisions according to the risk attitude and ability level

	Risk seeking	Risk averse	<i>p</i> – value
Stop after primary level			
High-ability	.05	.12	.0604
Medium-ability	.13	.27	.0069
Low-ability	.31	.49	.0098
Proportion of participants who select GEN			
High-ability	.82	.69	.2675
Medium-ability	.78	.56	.0872
Low-ability	.67	.67	1.0000

Notes: **Stop after primary level** is conditional to success at this level. We present values for all treatments, except for Choice-low, in the analysis of this variable. **Proportion of participants who select GEN** is conditional to success at compulsory level and decision to start the secondary level. We present values exclusively for the Choice treatment in the analysis of this variable, because it is the unique treatment in which all participants who succeed compulsory level make this decision. **Number of observations** deciding to stop or not after primary level: Risk seeker: 113 HA, 110 MA, 81 LA. Risk averse 131 HA, 114 MA, 109 LA. **Statistical test** Two-sample test of proportions.

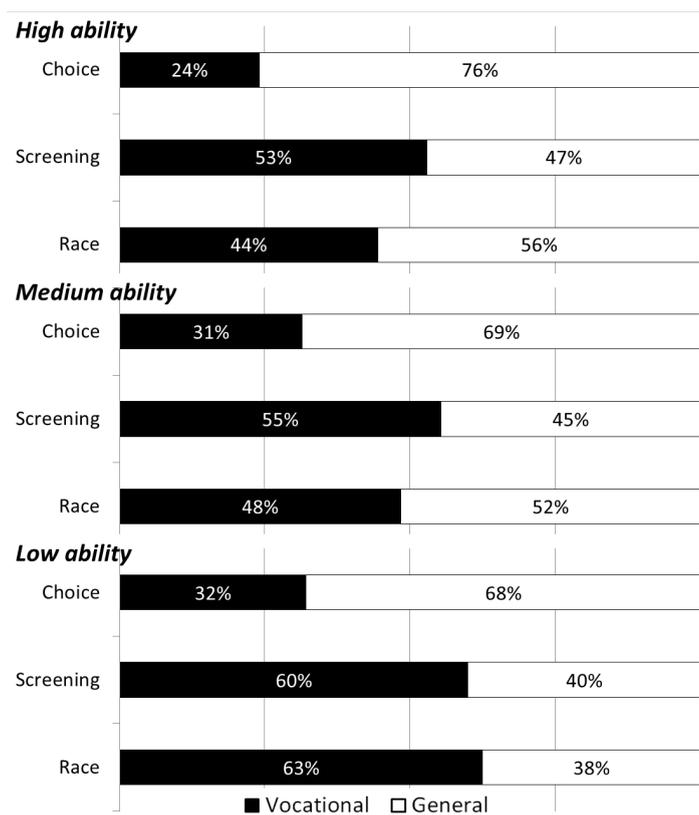
However, since the decision to start the post-compulsory level is not restricted to the Choice treatment⁹, the higher inequality observed in this treatment must be caused by the differential massive selection of the general track in this treatment in comparison with Screening and Race (see Figure 4.3), where LA have about 50%¹⁰ lower chances of success at secondary level¹¹

9. Moreover, Table 4.11 in the Appendix (page 212) shows that the decision to stop after level 1 is invariant to the school system (differences are insignificant at the 10% level).

10. For low ability individuals, chances of success at secondary level are 23% in the VOC and 12% in the GEN. For more information, ex-post chances of success are displayed in the Appendix to this Chapter (Table 4.14, page 217).

11. Indeed, the proportion of HA in the general track in this treatment is also higher than in others, but as chances of success for HA are similar in both tracks, this decision does not impact their performance in the experiment, increasing the inequality on achievements between low and high ability participants.

Figure 4.3 – Distribution of participants between the two possible tracks by ability level.



Note: participants that started the secondary level.

Choice is bad for equality when returns to further education are high because students are inclined to opt for the more difficult track (GEN) at secondary level when they can choose, either because they overestimate their chances or merely because the high returns more than compensate for low objective chances. Since potential returns do not depend here on ability, the choice of track does not differ much between ability levels but the pass rates obviously do. This effect should be dampened with Screening and Race because some students at least are forced to opt for the easier track.

5. Gender differences on educational achievement

Result 3: Men and women perform equally in the No-choice condition.

However, both Choice and Race are harmful to women but not to men.

Reproducing the description of ability and socioeconomic biases for gender, Table 4.6 and Figures 4.4 and 4.5 compare outcomes and attainment levels of men and women for all treatments. No-choice is the only treatment in which women and men have on average equal payoffs, outputs, and equivalent level achievements. In all other treatments, women perform worse than men on average. Not surprisingly, Race is the worst treatment for women who have been shown to shy away from competition (e.g. Gneezy, Niederle and Rustichini, 2003, Niederle and Vesterlund, 2008). Figure 4.4 shows that women earn only 63% of men's payoffs and solve 10% less anagrams. Furthermore, only 1% of women who started the experiment in the Race treatment reached the tertiary level (Figure 4.5).

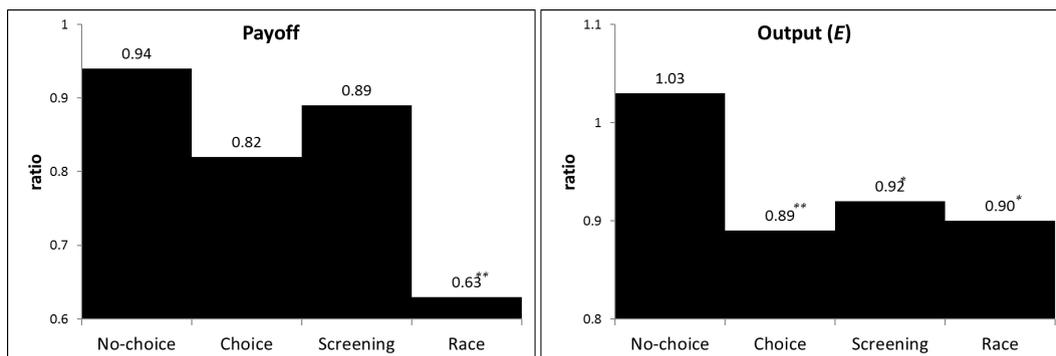
Table 4.6 – Variation of performance across sorting mechanisms by gender

	No-choice	Choice	Screening	Race
Women				
Payoff (€)	8.72 (7.83)	7.06 (6.13)	7.70 (6.95)	5.55 (4.22)
Total anagrams solved	55.56 (19.47)	51.16 (20.03)	54.02 (20.27)	45.65 (14.86)
Dropout rate at compulsory level	.15	.17	.21	.26
Dropout rate at secondary level	.62	.76	.66	.82
Attainment rate at secondary or tertiary levels	.26	.14	.20	.10
Men				
Payoff (€)	9.25 (7.75)	8.59 (7.93)	8.66 (7.41)	8.77** (8.12)
Total anagrams solved	53.81 (20.29)	57.63** (21.45)	58.69 (18.67)	50.98** (20.56)
Dropout rate at compulsory level	.10	.09*	.07***	.20
Dropout rate at secondary level	.64	.74	.68	.61**
Attainment rate at secondary or tertiary levels	.23	.19	.24	.25***

Notes: **Dropout rate at secondary level** is conditional on passing primary level and deciding to start the secondary. **Standard deviation** in parentheses. **Significance tests:** Payoff: two-sample Wilcoxon rank-sum (Mann-Whitney) test. Total anagrams solved: two-sample t-test with equal variances. Dropout and attainment rates: two-sample test of proportions. Gender differences within treatment: Sig. level: * 10%; ** 5%; ***1%; no symbol: not significant at 10% level. Significance tests are displayed in the Appendix to this Chapter (Table 4.6, page200) .

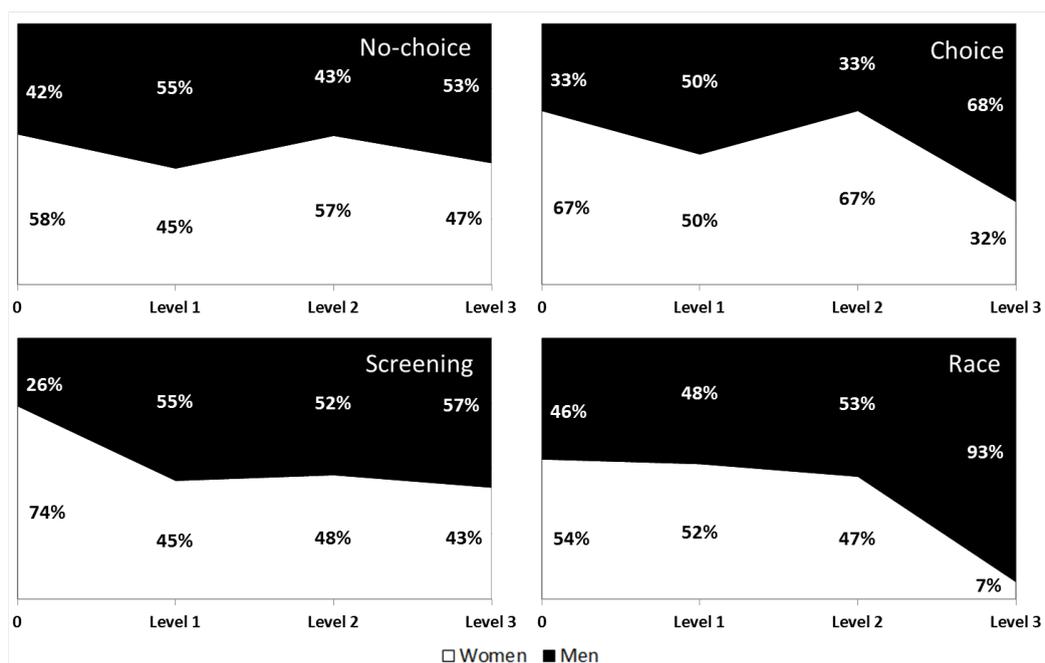
In the Choice and in the Screening treatments too, women present on average lower payoffs and lower outputs than men. Moreover, women fail more than men as soon as the primary level in these treatments. These observations are reminiscent of the gender

Figure 4.4 – Ratio of average payoffs and outputs (women/men)



Note: Gender differences within treatment: Significance level: * 10%; ** 5%; ***1%; no symbol means that it is not significant at 10% level (Wilcoxon rank-sum (Mann-Whitney) test).

Figure 4.5 – Proportion of genders at each level attained



Note: Level 1: succeeds the first level, does not start or fails the second level. Level 2: succeeds the second level, does not start or fails the third level.

gap found by Page, Lévy-Garboua and Montmarquette (2007) with a similar design in which they manipulated the aspiration level of participants. Women performed like men when the aspiration level was low but they performed worse than men when the

aspiration level was high. Similarly, during the primary level, the higher goal that participants should reach for being selected in the Screening treatment, or if they decide to continue in the Choice treatment, acts like an explicit aspiration level which is probably more salient than the implicit goal present in the No-choice condition. The perception of a high aspiration level seems to demotivate women and to motivate men, thus causing a gender gap. On reflection, the desire to win over a competitor may also be viewed as a high, and particularly salient, aspiration so that all our observations would find a common explanation. Accepting a difficult challenge is like entering a tournament, and men find this kind of situation more attractive than women.

6. The effect of financial incentives

The Choice-low treatment replicates exactly the Choice treatment with lower payoffs at the tertiary level. The mere change of returns to tertiary education impacts the comparison of the expected utilities of potential decisions, thus affecting all choices including those concerning lower levels of education. For instance, the proportion of participants quitting the task after succeeding the primary level is much higher for all ability groups in the Choice-low treatment relative to Choice (Table 4.7).

Table 4.7 – Variation of the frequency of quits after the primary level, by ability tercile

	Choice	Choice-low	<i>p-value</i>
High-ability	.11	.19	.2691
Medium-ability	.19	.43	.0072
Low-ability	.42	.60	.1048

Statistical test: Two-sample test of proportions.

Result 4: A rise in the returns to tertiary education is beneficial to high and medium-ability students but it is harmful to low-ability students. Table 4.8 immediately exhibits a striking result. By all standards, performance diminishes with diminishing incentives, but only for HAs and MAs. More surprisingly, LAs experience

rising performance under the same conditions. For instance, payoffs are cut by 24.5% and 12.5% for HAs and MAs respectively, whereas they increase by 22.3% for LAs.

Table 4.8 – Variation of outcomes with incentives, by ability tercile

	Choice	Choice-low	<i>p-value</i>
High ability			
Payoff (€)	11.37 (9.21)	8.59 (5.36)	.3132
Total anagrams solved	69.19 (21.06)	60.92 (17.96)	.0259
Dropout rate at compulsory level	.02	.02	.6890
Dropout rate at secondary level	.56	.60	.6268
Attainment rate at post-compulsory levels	.39	.31	.4166
Medium ability			
Payoff (€)	6.94 (6.04)	6.07 (3.61)	.5881
Total anagrams solved	54.22 (16.24)	46.04 (12.27)	.0025
Dropout rate at compulsory level	.07	.21	.0207
Dropout rate at secondary level	.84	.88	.6599
Attainment rate at secondary or tertiary levels	.12	.06	.2081
Low ability			
Payoff (€)	5.47 (4.13)	6.69 (3.95)	.0948
Total anagrams solved	40.87 (15.67)	40.63 (13.38)	.9310
Dropout rate at compulsory level	.31	.27	.6618
Dropout rate at secondary level	.96	.75	.0352
Attainment rate at secondary or tertiary levels	.01	.07	.1052

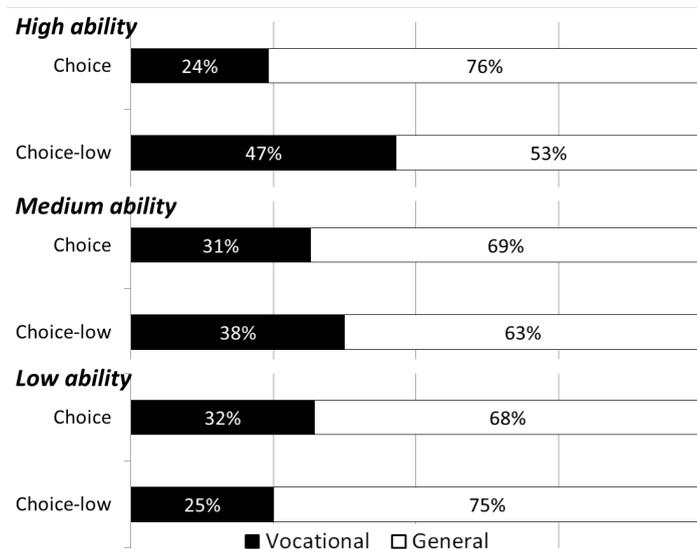
Notes: **Dropout rate at secondary level** is conditional on passing primary level and deciding to start the secondary. **Attainment rate at post-compulsory levels** refers to participants who succeed secondary or tertiary levels. **Statistical tests:** Payoffs: two-sample Wilcoxon rank-sum (Mann-Whitney) test. Total anagrams solved: two-sample t-test with equal variances. Level attained and decisions: two-sample test of proportions. **Standard deviation** in parenthesis.

A rise in returns to tertiary education hurts low-ability students because increasing the returns to higher education pushes low-ability students to engage in secondary education and the general track instead of stopping after primary level or going for vocational studies, thus raising their probability of failure and causing regret. In the higher ability terciles, raising the higher level's payoff exerts the same attraction towards tertiary education but produces an opposite effect with lower failure rates in post-compulsory levels. However, a significant share of MAs -likely among the risk-averse- lose their motivation under low incentives to the point of failing the primary level. Typically, MAs seem to be essentially motivation-driven: under low incentives, they drop out at secondary and tertiary levels *more* than LAs. And there is not a

single medium-ability participant succeeding at the higher level in our experiment.

We observe also in Figure 4.6 a strong impact of incentives in the self-selection of the track: the proportion of HAs choosing the safer and probably shorter vocational track almost doubles if the returns to tertiary education diminish. The corresponding effect is much smaller in the other ability groups, probably revealing an overconfidence of lower-ability subjects relative to the high-ability group.

Figure 4.6 – Distribution of participants between the two possible tracks.



Note: Participants that started the middle level.

Result 5: A rise in the returns to tertiary education increases the socioeconomic bias of school systems. The socioeconomic-biasing effect of high returns to tertiary education is highlighted by Figures 4.7 and 4.8. LAs are obviously better-off under low returns relative to HAs, which is a direct consequence of Result 4. The ratio between low and high ability payments and outputs is much higher in the Choice-low. The problem of Choice-low concerns the level attained: there is not a single MA or LA participant reaching the tertiary level in our experiment. The lower-ability groups aim at less ambitious goals as it doesn't pay their (high) effort to reach the tertiary level.

They are happy with safer investments at lower levels and their cautious behavior has the virtue of enabling them to reach secondary education by the safer track. At least we have a better ratio of low-ability participants reaching the secondary level (the same observed in other treatments, except for Choice).

Figure 4.7 – Ratio of average payoffs and outputs (low/high ability)

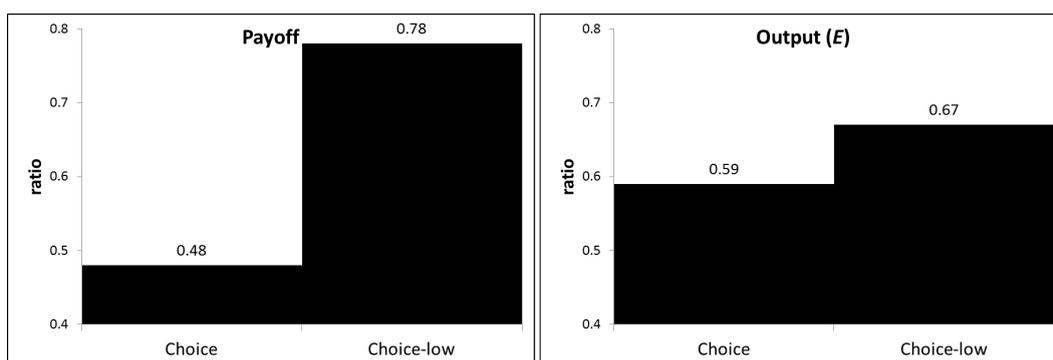
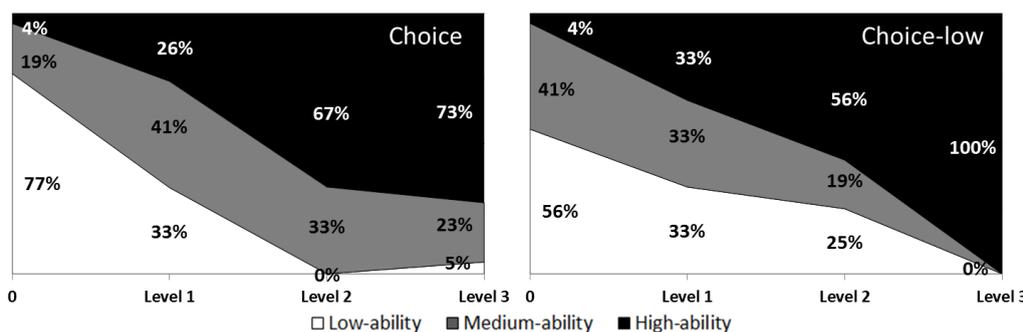


Figure 4.8 – Participation of the three ability terciles at each level attained in the task.



Note: Level 1: succeeds the first level, does not start or fails the second level. Level 2: succeeds the second level, does not start or fails the third level.

Result 6: A rise in the returns to tertiary education increases the gender bias of school systems. Table 4.9 and Figures 4.9 and 4.10 depict the gender-biasing effects of incentives. Lower returns to tertiary education negatively affects the performance of men and women alike. As we discussed earlier, women are at a relative

disadvantage in Choice treatments and the reduction of incentives does not change that. The main difference seems to be that men are more driven by motivation and incentives than women since performance (in terms of output and success) is more sensitive to a change of incentives for men than for women (see Table 4.9). Hence, as shown by Figures 4.9 and 4.10, a rise in the returns to tertiary education would tend to increase the gender bias of the school system in terms of educational performance. However, this conclusion doesn't extend to payoffs, perhaps because men adopt riskier choices on average.

Table 4.9 – Variation of performance with incentives, by gender

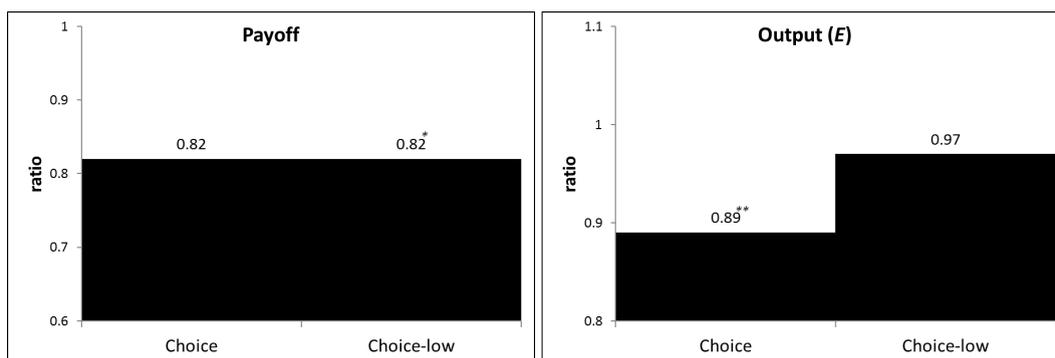
	Choice	Choice-low	<i>p-value</i>
Women			
Payoff (€)	7.06 (6.13)	6.42 (4.09)	.9264
Total anagrams solved	51.16 (20.03)	48.44 (16.87)	.3281
Dropout rate at compulsory level	.17	.18	.8310
Dropout rate at secondary level	.76	.83	.4707
Attainment rate at secondary or tertiary levels	.14	.09	.2281
Men			
Payoff (€)	8.59 (7.75)	7.83* (4.76)	.5695
Total anagrams solved	57.63** (21.45)	49.89 (17.23)	.0093
Dropout rate at compulsory level	.09*	.15	.2340
Dropout rate at secondary level	.74	.60**	.1413
Attainment rate at secondary or tertiary levels	.19	.21**	.7644

Notes: **Dropout rate at secondary level** is conditional on passing primary level and deciding to start the secondary. **Attainment rate at post-compulsory levels** refers to participants who succeed secondary or tertiary levels. **Statistical tests:** Payoffs: two-sample Wilcoxon rank-sum (Mann-Whitney) test. Total anagrams solved: two-sample t-test with equal variances. Level attained and decisions: two-sample test of proportions. **Gender differences within treatment:** Sig. level: * 10%; ** 5%; ***1%; no symbol: not significant at 10% level. **Gender differences between treatment:** displayed in the last column. **Standard deviation** in parenthesis.

7. Conclusion

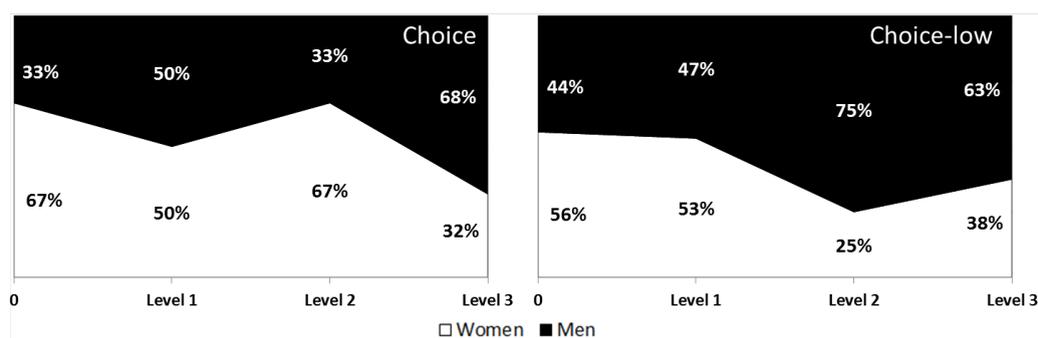
How do different school systems and school returns affect differently ability groups, genders, and social groups, thus causing substantial differences in social and gender bias among developed countries and periods? We have proposed an experimental setting with five treatments investigating four different mechanisms to assign students to

Figure 4.9 – Ratio of average payoffs and outputs (women/men)



Note: gender differences within treatment: Sig. level: * 10%; ** 5%; ***1%; no symbol means that it is not significant at 10% level (two-sample Wilcoxon rank-sum (Mann-Whitney) test).

Figure 4.10 – Proportion of participants at each level attained



educational curricula according to their capacity to produce at school: self-selection of further studies with a single track (no-choice of track), self-selection of further studies with the choice of track, screening by ability and early *numerus clausus* competition.

The impact of mechanisms is different according to participants' ability. The performance of high-ability individuals is very much alike in all treatments. The only exception is the early competition (Race) where high-abilities perform worse than in other treatments. As at the compulsory level participants are screened in the speed of resolution of anagrams to be able to select the track for post-compulsory levels, they minimize the time of resolution of anagrams instead of maximizing the number of

solved anagrams.

Medium-ability participants are the most affected by treatments. Competition is the worst treatment for them because this mechanism discourages medium-ability individuals and eliminates them too early from the experiment. Consequently medium-ability individuals perform like low-ability individuals at the compulsory level. Self-selection leads to the highest level of failure of medium-ability participants at secondary level, showing that the higher is the ratio of individuals selecting their tracks for secondary and tertiary levels, the lower is the efficiency. The performance of medium-ability individuals is similar for random allocation and screening.

Self-selection of track is specially unfair for low-ability participants because of the rate of failure at secondary level. The chances of success at post-compulsory levels are weak for these participants, but high incentives for higher education force them to try start the post-compulsory level and they fail. They maximize their ex-ante expected utility, but it does not increase their chances of success. This may explain why school systems contribute to socioeconomic biases: the regret caused by educational decisions to low-ability (or low socio economic status) individuals increases the inequality on payoffs, educational output and level attained between low and high ability students.

Additionally, early competition is also a problem to low ability participants, because as for medium-ability and high-ability individuals, their performance is disturbed and they are eliminated too early from the experiment. Low-ability individuals perform better in the random allocation and screening, even if these treatments are not able to decrease the regret observed for less able individuals in the experiment.

We also investigate the impact of payoffs for higher levels of education (wage premium) comparing Choice to the Choice-low condition. The main result is that a rise in the returns to tertiary education is beneficial to high and medium-ability students but it is harmful to low-ability students. A consequence of this result is that returns to tertiary education increases the socioeconomic bias of school systems. This result may

explain differences in social and gender bias among developed countries and periods.

The investigation of *social* inequalities on educational achievement brings different conclusions about sorting mechanisms. Selection by ability decreases inequalities as early competition and screening promote the most equitable outcomes for low and high ability participants. Note that the "equity" promoted by early competition is due mainly to the inefficiency of this treatment for high-ability participants. Unlike self-selection increases inequalities, but inequality decreases if the wage premium for higher levels decreases. The random allocation is the only mechanism that is fair for gender differences.

Several clear policy implications emerge from our results in the challenge to design a system that is efficient and equitable at same time. First, as there is no big impact for high ability individuals, policies must focus on medium and low-ability individuals and should increase their motivation and productivity without discouraging high-ability students. Screening seems to be the most balanced mechanism to track students by ability, the challenge is to set fair grades (thresholds) that encourage motivated low and medium ability students to reach higher levels of education, without discouraging the less motivated to complete the primary level. An equitable system should improve the outcomes of less able individuals, without prejudice for the more able.

Second, the efficiency of the No-choice treatment is directly linked to statistical issues, as we have participants of all levels of ability equally distributed on the vocational and general tracks. We do not advocate that this is the best solution for educational institutes, as we believe that students must choose their preferred track, but we cannot neglect that educational institutions must also support students in the orientation choices, since the self-selection has an important weight in the educational decisions even if institutes filter students by ability level. Making educational choices is hard, since when deciding the student does not know (perfectly) her ability and the difficulty that she will face in the future. Policymakers must design not only efficient sorting

mechanisms, but help less able students to understand the difficulties they will face in the future, reducing the regret caused by education. It seems crucial to evaluate on the gateway not only performance but also intrinsic motivations and other non-cognitive abilities, especially for less able individuals.

Third, early competition must be avoided. Under competition students worry about others performance leaving aside the own educational output, which should be the major goal of education. Moreover competition discourages low and medium ability students. It is more equitable than self-selection and random allocation, but much more inefficient than these mechanisms.

External validity and limitations. Our experimental design has two main limitations when compared to a real school system: *(i)* we don't observe differences in opportunities, *(ii)* impatience and discount rates are not observed because schooling time is compressed in our experiment. However, we don't have the ambition of reproducing the school system with all its characteristics and complexities in the laboratory. First because we aim at identifying a specific causal relationship, second because reproducing perfectly a school system is impossible, and finally even in the case where it is possible, because we would use real data instead since they are more natural. According to [Camerer \(2015\)](#), reproducing all characteristics of the environment studied is not a primary concern in a typical experiment, since experimental economics aims at establishing a general theory linking economic factors such as incentives, rules and norms to decisions and behavior. However, our findings are coherent to literature, suggesting that our experimental framework is externally valid. For instance, [Hanushek and Wößmann \(2006\)](#) found that early tracking increases educational inequality and reduces the mean performance, very similar to results observed in the Choice treatment. Another example comes from [Arenas and Malgouyresy \(2017\)](#) recent result, authors show that bad economic times are good for social mobility. They argue that the opportunity cost of education is lower during recessions, favoring the educational attainment

for children of low-income families. We found the same result experimentally in the Choice-low condition, ruling out the assumption of difference in opportunities.

A. Appendices

A.1. Tables and figures

Table 4.10 – Main outcomes by ability level: statistical tests for Table 4.3

<i>p</i> – value	No-Choice			Choice		Screening	Test
	Choice	Screening	Race	Screening	Race	Race	
High ability							
Payoff (€)	0.7917	0.5099	0.0556	0.6886	0.0800	0.1886	(i)
Total anagrams solved	0.4356	0.8983	0.0379	0.3696	0.0058	0.0498	(ii)
Dropout rate at compulsory level	0.6233	0.5628	0.8735	0.3094	0.5298	0.6890	(iii)
Dropout rate at secondary level	0.6069	0.5183	0.0313	0.8927	0.1109	0.1500	(iii)
Attainment rate at post-secondary levels	0.4715	0.2715	0.0372	0.7100	0.1755	0.3217	(iii)
Medium ability							
Payoff (€)	0.5036	0.9510	0.1425	0.5541	0.2918	0.1418	(i)
Total anagrams solved	0.7178	0.1332	0.0431	0.2020	0.0104	0.0005	(ii)
Dropout rate at compulsory level	0.1386	0.5860	0.0331	0.3662	0.0004	0.0105	(iii)
Dropout rate at secondary level	0.0561	0.9430	0.7936	0.0649	0.0439	0.7435	(iii)
Attainment rate at secondary or tertiary levels	0.1348	0.7385	0.6305	0.0724	0.3350	0.4294	(iii)
Low ability							
Payoff (€)	0.0613	0.3875	0.1244	0.3403	0.9318	0.4579	(i)
Total anagrams solved	0.1645	0.5730	0.0473	0.0722	0.5593	0.0207	(ii)
Dropout rate at compulsory level	0.1552	0.4502	0.0609	0.5105	0.6205	0.2589	(iii)
Dropout rate at secondary level level	0.1113	1.0000	0.6953	0.1113	0.0581	0.6953	(iii)
Attainment rate at secondary or tertiary levels level	0.0986	0.9418	0.8179	0.0876	0.0668	0.8752	(iii)

Statistical tests: (i) Two-sample Wilcoxon rank-sum (Mann-Whitney) test, (ii) Two-sample t-test with equal variances, (iii) Two-sample test of proportions.

Table 4.11 – Proportion of participants who decide to stop after succeed the compulsory level.

	No-choice	Choice	Screening	Race
High ability	6%	11%	14%	5%
Medium ability	22%	19%	15%	26%
Low ability	45%	42%	39%	38%

Note: Statistical tests are displayed in table 4.12.

Table 4.12 – Proportion of participants who decide to stop after succeed the compulsory level: statistical tests (Table 4.11)

$p - value$	No-Choice			Choice		Screening	Choice-low
	Choice	Screening	Race	Screening	Race	Race	Choice
High ability	0.2689	0.1492	0.8655	0.7311	0.2279	0.1290	0.2712
Medium ability	0.6858	0.3048	0.6303	0.5046	0.3851	0.1543	0.0074
Low ability	0.7697	0.4934	0.5016	0.7078	0.7023	0.9762	0.1068

Two-sample test of proportions

Table 4.13 – Main outcomes by gender: statistical tests (Table 4.6)

$p - value$	No-Choice			Choice		Screening	Test
	Choice	Screening	Race	Screening	Race	Race	
Women							
Payoff (€)	0.2034	0.3368	0.0060	0.7957	0.0945	0.0794	(i)
Total anagrams solved	0.1142	0.5899	0.0002	0.3200	0.0357	0.0020	(ii)
Dropout rate at compulsory level	0.6773	0.2697	0.0769	0.4788	0.1638	0.4952	(iii)
Dropout rate at secondary level	0.0755	0.6638	0.0200	0.2032	0.4798	0.0633	(iii)
Attainment rate at secondary or tertiary levels	0.0357	0.3205	0.0062	0.2807	0.4137	0.0715	(iii)
Men							
Payoff (€)	0.2993	0.6960	0.2350	0.5315	0.6772	0.3723	(i)
Total anagrams solved	0.1893	0.0770	0.3273	0.7158	0.0295	0.0076	(ii)
Dropout rate at compulsory level	0.7901	0.4671	0.0421	0.6472	0.0261	0.0088	(iii)
Dropout rate at secondary level	0.2296	0.6349	0.7020	0.4633	0.1244	0.4011	(iii)
Attainment rate at secondary or tertiary levels	0.4871	0.8917	0.7930	0.4183	0.3535	0.9015	(iii)

Statistical tests: (i) Two-sample Wilcoxon rank-sum (Mann-Whitney) test, (ii) Two-sample t-test with equal variances, (iii) Two-sample test of proportions.

Figure 4.11 – Low vs. medium-ability

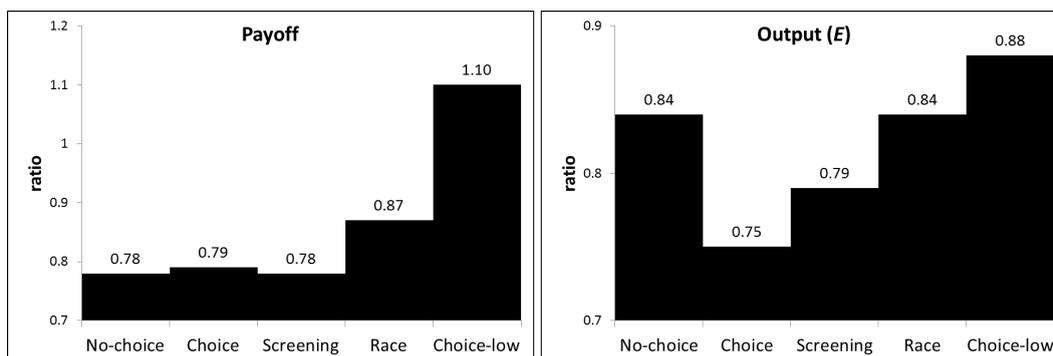
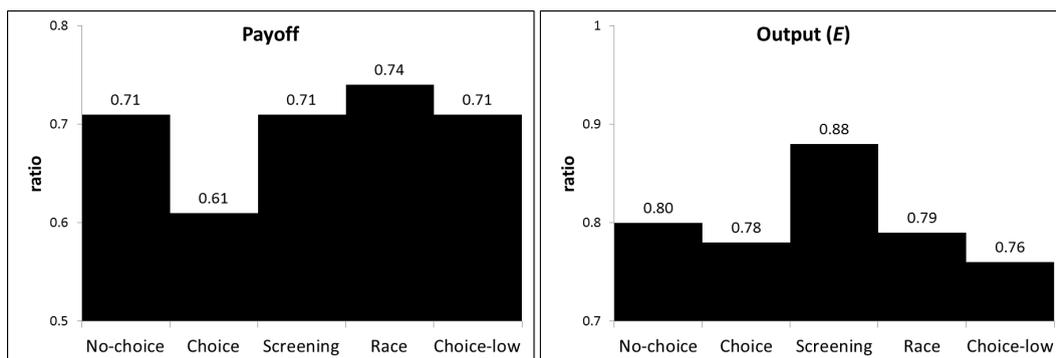


Figure 4.12 – Medium vs. high-ability



A.2. Model of rational decisions

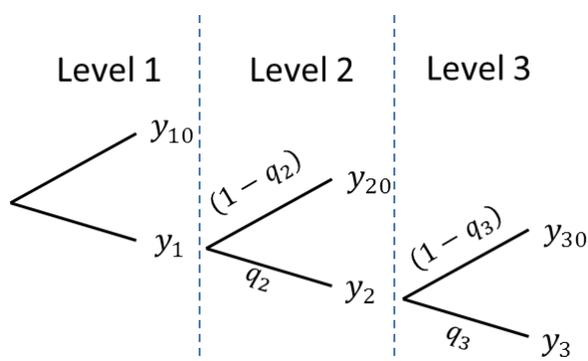
We present a simple model of rational decisions, based on the outcomes of the experiment and participants' ability level¹², in order to identify the optimal expected decision for each ability group in the experience.

Each participant can make at most three different decisions, depending on the treatment assigned, the performance and achievements:

- In all treatments, if one succeeds the first level, she can decide to stop and receive her payment, or to continue the second level;
- Those who decide to continue (and can opt for the track according to treatment), must select the preferred path: VOC or GEN;
- Finally, those who succeed the second level are asked to start the third level, or to quit after the second level (all treatments).

General conditions: outcomes and chances of success The experiment can be summarized and represented by figure 4.13:

Figure 4.13 – Decision tree: probabilities are conditional on prior success.



Where:

12. Ability is measured by the number of anagrams solved per minute in the four first rounds of the experience.

- y_{10} : outcome if fails the first level.
- y_1 : outcome if succeeds the first level and doesn't start level 2.
- y_{20} : outcome if decides to start the second level, but fails at this level. Note that: $y_{10} < y_{20} < y_1$.
- y_2 : outcome if succeeds the level 2 and quit the experiment after it.
- y_{30} : outcome if starts and fails the third level. Note that: $y_{30} < y_2$.
- y_3 : outcome if succeeds level 3.
- q_2 : probability to clear the second level. It is important to highlight that this probability is supposed to be different depending on the selected/imposed path. As GEN and VOC have a different number of anagrams to be solved at the second level. In the model, the different probabilities for different treatments are identified with an index "G" when GEN and "V" when VOC.
- q_3 : probability to clear the third level. It is important to observe that (theoretically) in the GEN condition $q_{(3|2)}^G = 1$, because the number of anagrams to be solved is exactly the same in the second and third levels in this track; in other words, if the track is GEN, one is able to complete level 3 if he or she is also able to clear the second level ¹³.
- $q_{(3|2)}$: is the probability to clear the third level conditional to clear level 2. As the chances of success are the same in the second and third levels in the GEN condition, this variable is not useful for this track. Note that: $q_{(3|2)} = \frac{q_3}{q_2}$, $q_{(3|2)}^G = 1$, and $q_{(3|2)}^V < 1$.

Decisions after level 1 Individuals who succeeded the first level have 4 possible outcomes when considering the Choice treatment. A rational agent is supposed to

13. When assuming that $q_{(3|2)}^G = 1$, we don't consider the potential fatigue of participants in this long task. Ex-post results (for No-choice) shows that for HA $q_2^G = 0.79q_3^G$, for MA $q_2^G = q_3^G$, and LA $q_2^G = 0.5q_3^G$. Keeping our assumption doesn't change the general results from our model, anyway we consider both $q_{(3|2)}^G = 1$ and $q_{(3|2)}^G \neq 1$ in the analysis.

select the option that maximizes the expected utility (net of the cost of effort if they continue) comparing the 4 possibilities:

- [1] Stop after level 1:

$$EU_{11} = u(y_1)$$

- [2] Select VOC and stop after level 2:

$$EU_{12}^V = (1 - q_2^V) \cdot u(y_{20}) + q_2^V \cdot u(y_2)$$

- [3] Select VOC and go to level 3:

$$EU_{13}^V = (1 - q_2^V) \cdot u(y_{20}) + q_2^V \cdot \left[q_{(3|2)}^V \cdot u(y_3) + (1 - q_{(3|2)}^V) \cdot u(y_{30}) \right]$$

- [4] Select GEN and go to level 3¹⁴:

$$EU_{13}^G = (1 - q_2^G) \cdot u(y_{20}) + q_3^G \cdot u(y_3)$$

Table 4.14 present the chances of success for GEN and VOC for a given ability level in the No-choice treatment. We consider the chances of success in the No-choice to avoid the effect of other treatments in the chances of success for higher levels, anyway results are not qualitatively different when considering the chances of success for all treatments.

Table 4.14 – Ex-post chances of success for No-choice participants

Ability level:	High		Medium		Low	
	VOC	GEN	VOC	GEN	VOC	GEN
q_2	57%	42%	36%	29%	23%	12%
q_3	30%	33%	16%	29%	0%	6%

Table 4.15 present the EU for each level and track by ability level. The conclusion is that Low-ability individuals, independently of the risk behavior, should stop after the the first level. Medium ability individual should start the GEN track at secondary level, with the exception of risk averse individuals whose should stop after level 1.

High ability individuals should always start the secondary level as their EU for the tertiary level, for both tracks, are always higher than the status quo (stop after level

14. Assuming outcomes for the third level, as the chances of success are the same for second and third levels in the GEN (same challenge for both levels).

1). GEN seems to be a better alternative, independently of the risk behavior, but VOC can also be selected by risk averse individuals. Conclusions are slightly different when applying the same reasoning to the Choice-low payoff matrix: HA risk seekers should opt for VOC and stop after secondary level or go to the third level in the GEN; MA risk seekers that should opt for GEN, all other participants are supposed to stop after level 1.

Table 4.15 – Expected utility for different ability levels and risk profiles

	Risk neutral	Risk seeker (y^2)	Risk averse (\sqrt{y})
High-ability			
EU_{l1}	10	100	3.16
EU_{l2}^V	9.70	118.60	2.99
EU_{l3}^V	12.49	242.35	3.29
EU_{l3}^G	13.24	293.20	3.30
Medium-ability			
EU_{l1}	10	100	3.16
EU_{l2}^V	7.60	80.80	2.63
EU_{l3}^V	8.92	142.60	2.76
EU_{l3}^G	10.38	207.40	2.90
Low-ability			
EU_{l1}	10	100	3.16
EU_{l2}^V	6.30	57.40	2.40
EU_{l3}^V	5.61	40.15	2.30
EU_{l3}^G	6.64	95.20	2.37

Notes: **These numbers neglect the cost of effort (time).** The latter might be introduced under the quadratic cost function with an equal value of the cost parameter for all ability levels (and imputing to each group the average time devoted to the task at various levels and tracks). **If we relax our theoretical assumption** $q_{(3|2)}^G = 1$ and use the ex-post chances of success: (i) the expected value for risk neutral HA at third level in the GEN drops to 11.89, in this case HA neutral to risk should opt for VOC; (ii) the expected value for risk seekers HA at third level in the GEN drops to 243.25, in this case HA risk seekers should opt for VOC or GEN; (iii) the expected value for risk averse HA at third level in the GEN drops to 3.15 in this case HA averse to risk should opt for VOC or to stop.

Table 4.16 – Decisions according to the risk attitude

	Risk seeker	Risk averse	<i>p</i> – value
Stop after primary level			
High-ability	.05	.12	.0604
Medium-ability	.13	.27	.0069
Low-ability	.31	.49	.0098
Proportion of participants who select GEN			
High-ability	.82	.69	.2675
Medium-ability	.78	.56	.0872
Low-ability	.67	.67	1.0000

Notes: **Stop after primary level** is conditional to success at this level. We present values for all treatments, except choice-low, in this analysis. **Proportion of participants who select GEN** is conditional to success at this level. We present values for the Choice treatment, because it is the only treatment where all participants make this decision. **Statistical test**: two-sample test of proportions.

Table 4.17 – Expected utility for different ability levels and risk profiles (low incentives)

	Risk neutral	Risk seeker (y^2)	Risk averse (\sqrt{y})
High-ability			
EU_{l1}	10	100	3.16
EU_{l2}^V	9.70	118.60	2.99
EU_{l3}^V	9.28	121.36	2.90
EU_{l3}^G	9.88	145.36	2.94
Medium-ability			
EU_{l1}	10	100	3.16
EU_{l2}^V	7.60	80.80	2.63
EU_{l3}^V	7.04	74.88	2.52
EU_{l3}^G	8.06	105.32	2.65
Low-ability			
EU_{l1}	10	100	3.16
EU_{l2}^V	6.30	57.40	2.40
EU_{l3}^V	4.92	27.04	2.19
EU_{l3}^G	5.68	52.96	2.27

Notes: **These numbers neglect the cost of effort (time)**. The latter might be introduced under the quadratic cost function with an equal value of the cost parameter for all ability levels (and imputing to each group the average time devoted to the task at various levels and tracks). **If we relax our theoretical assumption** $q_{(3|2)}^G = 1$ and use the ex-post chances of success the expected value for risk neutral HA at third level in the GEN drops to 121.96, in this case HA neutral to risk should opt for VOC.

Table 4.18 – Decisions in the choice-low treatment according to the risk attitude

	Risk seeker	Risk averse	<i>p</i> – value
Stop after primary level			
High-ability	.10	.29	.0813
Medium-ability	.41	.44	.8560
Low-ability	.47	.68	.1824
Proportion of participants who select GEN			
High-ability	.58	.52	.6666
Medium-ability	.64	.41	.2001
Low-ability	.47	.14	.1317

Notes: **Stop after primary level** is conditional to success at this level. We present values for all treatments, except choice-low, in this analysis. **Proportion of participants who select GEN** is conditional to success at this level. We present values for the Choice treatment, because it is the only treatment where all participants make this decision. **Statistical test**: two-sample test of proportions.

GENERAL CONCLUSION

This thesis comprises four essays based on behavioral and experimental economics approaches, with two main ambitions. Our first ambition is to study two noncognitive skills, namely self-confidence and creativity. We study the determinants of self-confidence, and the impact of creative potential on economic outcomes. The second ambition is to study how school systems impact educational decisions, educational outcomes and intergenerational mobility, where noncognitive skills play an important role, specially self-confidence and motivation. This general conclusion provides an overarching view of the contributions of this thesis and presents perspectives of research.

1. Noncognitive abilities are important for life success

We contribute to the growing economic literature putting in perspective the importance of noncognitive skills to explain economic behavior. The first contribution to this field of research concerns the determinants of self-confidence when estimating the subjective chances of success before completion of a task (Chapter 1). We present a new intuitive-Bayesian model, which was tested on experimental data. According to our model, estimation biases are likely to persist since the Bayesian aggregation of past information consolidates the accumulation of errors, and the perception of illusory signals generates conservatism and under-reaction to events. Taken together these two features of our model may explain why intuitive Bayesian make systematically wrong and costly predictions of their own performance. In our experiment analog to the popular double-or-quits game, the prevalence of miscalibration among doublers is substantial (41%) and its cost in terms of failures is massive: 91% of overconfident participants failed. This rate of failure is 39% larger than for able and calibrated individuals. This findings suggests that overconfident players, chiefs, entrepreneurs,

traders, or students are inclined to take excessive risks; they are unable to stop at the right time and eventually fail more than well calibrated persons.

In Chapter 2 we present an exploration into a new field for economists: the creative potential. Despite the incontestable importance of this variable, the economic literature on creativity is scarce. This chapter shows the importance of creative potential on individuals economic behavior, suggesting that creativity would be a valuable addition to common personality traits and IQ for predicting school achievements and future earnings. In a laboratory experiment, we measure the creative potential of a teenagers group and relate creative potential scores to the performance in real effort tasks and school grades. Our measures of creative potential, taken from [Lubart, Besançon and Barbot \(Evaluation du Potentiel Créatif, 2011\)](#), allow us to disentangle two aspects in the production of creative work: the divergent-exploratory and the convergent-integrative processes. We argue that from an economic perspective, creative behavior must be judged by individual's propensity to innovate in production (and consumption) activities, distinguishing two types of economic innovators: *researchers* (the ability to find new solutions to existing problems) and *entrepreneurs* (the ability to capture unexpected rents). We find that divergent *thinkers* are more productive than others in exploration activities, such result means that the psychological assessment of the divergent-exploratory aspect of creativity correlates with the researcher economic type of economic innovator. Our measures of creative potential didn't identify the entrepreneur type of innovator, thus the ability to capture unexpected rents may be related to other individual characteristics such as risk preferences and tenacity but not to creative potential as we expected. Moreover, we find that: (i) integrative *thinkers* are more cooperative when working in pairs, (ii) creative individuals are no more productive than others in repetitive tasks, and importantly, (iii) both measures of creativity are positively correlated to grades on Maths, French and general grade for 10th graders. The later results corroborates with the hypothesis that creative individuals

are intelligent.

This preliminary exploration brings evidences that creative potential is important even in day-by-day activities. Further research must address in what extent these evidences are related to creativity or to subprocesses involved in the two aspects of creative production (divergent-exploratory and convergent-integrative), such as flexibility, divergent thinking, selective encoding, associative thinking and selective comparison. Another important question of research derives from evidences that creative individuals may have different behavior at work: we show that they perform differently some tasks and may have different satisfaction - why not different motivation - for different jobs. Further research must investigate the extent in what creative behavior is required for different jobs, and how to match creative individuals to these jobs, increasing productivity and job satisfaction. Indeed, it was one of the ambitions of the present study. Unfortunately our scores of creativity are not correlated to the Holland's (1966) vocational typology used in our study, suggesting that this question must be addressed using a different methodology. Besides, advances in the economic literature depends on the development of simple, easily replicable tasks, like our buttons task, to identify pure types of economic creative behavior and make use of the resulting creativity score to test the impact of creativity in various contexts. Psychologists can greatly help economists to accomplish this objective.

2. On the use of experimental economics to investigate school systems

Another contribution of this thesis lies in the proposition of an experimental framework in order to investigate school systems, educational decisions and educational outputs. Our design can serve as a useful tool to provide clues for research, and even to test public policies in a cost-effective and non-impacting way for society. We reproduce

by means of an incentivized lab experiment the actual *minimal* school system¹, where we can discriminate differences in curricula, differences in payoffs, the choice of educational track and the performance level to qualify for a certain curriculum. Besides the comparison of different mechanisms to select students, our design allows us to observe participants' behavior when they are asked to make decisions based on their own ability, in a scenario of uncertainty like educational decisions.

There are several empirical limitations to study a given institutional context and/or make international comparisons as it is almost impossible to isolate the investigated effect maintaining everything else constant. The use of an experimental framework is a good alternative to study educational institutions, even if it is not possible to capture all the elements of an educational system under a controlled environment. The lab experiment allows us to measure variables like performance, ability, self-confidence and effort (time) that would be difficult to observe precisely in surveys. Experimental results allow an easy and valid comparison of the overall performance of a minimal school system under various sorting mechanisms, thus facilitating the identification of the efficient design that is, the educational output-maximizing design conditional on the ability distribution.

Our experimental design has two main limitations when compared to a real school system: (i) we don't observe differences in opportunities, (ii) impatience and discount rates are not observed because schooling time is compressed in our experiment. However, we don't have the ambition of reproducing the school system with all its characteristics and complexities in the laboratory. First because we aim at identifying a specific causal relationship, second because reproducing perfectly a school system is impossible, and finally even in the case where it is possible, because we would use real data instead since they are more natural. According to [Camerer \(2015\)](#), reproducing all characteristics of the environment studied is not a primary concern in a typical

1. We describe the characteristics of the minimal school system in the introduction to Part II (page 126).

experiment, since experimental economics aims at establishing a general theory linking economic factors such as incentives, rules and norms to decisions and behavior. However, our findings are coherent to literature (for instance: [Hanushek and Wößmann, 2006](#), [Filippin and Paccagnella, 2012](#), [Goux, Gurgand and Maurin, 2016](#), [Arenas and Malgouyresy, 2017](#))², suggesting that our experimental framework is externally valid.

3. Understanding the efficiency of school systems and inequalities on educational attainment

Our experimental setting allowed us to investigate and compare in Chapter 3 four different mechanisms to assign students to educational curricula according to their capacity to produce at school³: no choice of curriculum, self selection of track, early competition and screening. We found that no choice of curriculum and screening are the most efficient mechanisms. Early competition is the mechanism promoting less effort and consequently the highest level of failure in the first level of our experiment. Under competition students worry about others' performance leaving aside the own educational output, which should be the major goal of education. The problem of self-selection of track is that it promotes the highest level of failure at secondary level when the economic returns to school are high. We observe that participants selected massively the more difficult track (general) and failed. This decision may be ex-ante efficient, because participants maximize their EU, but ex-post inefficient because it decreases by 24% the wealth generated in the Choice treatment. Our evidences suggest that the inefficiency of the system derives from two main reasons: (i) if students have an imperfect knowledge of their own ability, and/or if there is a lack of discrimination between the two tracks (Vocational and General), they are inclined to opt for the more difficult track and fail; (ii) the higher is the wage premium for tertiary education the

2. We provide a comparison of these studies to our results in the conclusions for chapters 3 and 4.

3. Under two incentives structures.

higher is the ex-ante expected utility, what boosts the chances to try higher levels of education, but does not increase the probability of success, raising the ex-post inefficiency (higher level of failures, dropouts and regret).

Additionally, our investigation of inequalities on educational attainment in Chapter 4 shows that selection by ability decreases inequalities as early competition and screening promote the most equitable outcomes for low and high ability participants. The "equity" promoted by early competition is due mainly to the inefficiency of this treatment for high-ability participants. Unlike self-selection increases inequalities, but inequality decreases if the wage premium for higher levels decreases. The random allocation is the only mechanism that is fair for gender differences. We argue that an equitable system should improve the outcomes of less able individuals, without prejudice for the more able. Thus when comparing results presented in chapters 3 and 4, both self-selection of track and early competition don't present a good balance between efficiency and equity.

Several clear policy implications emerge from our results in the challenge to design a system that is efficient and equitable at same time. First, screening students by ability seems to be the most efficient and equitable mechanism to track students by ability. Second, policymakers must design not only efficient sorting mechanisms, but help students understand the difficulties they will face in the future, reducing the dropout and failures on higher levels of education. Third, early competition must be avoided. Under competition students worry about others' performance leaving aside the own educational output, which should be the major goal of education. It is more equitable than self-selection and random allocation, but much more inefficient than these mechanisms.

A perspective of research is to evaluate new sorting mechanisms and educational policies in our experimental framework. For instance, by introducing the decision of track during the compulsory level, thus simulating schools systems with early tracking.

The introduction of a treatment where participants are informed about dropout rates in the secondary and tertiary levels may give important cues to the second educational policy implication discussed in the previous paragraph. Additionally, the introduction of new measures such as time inconsistency and time preferences may give important cues about the decision process for selecting further studies and selecting the educational track. Still, a test of vocabulary, as a measure of crystallized intelligence, would be a useful and important control variable.

4. Education policy agenda must address noncognitive abilities

When summarizing the main results of this thesis, we observe that noncognitive abilities are essential to economic behavior and economic outputs - specially important to educational achievements. Self-confidence studied in Chapter 1 influences school decisions observed in chapters 3 and 4. Chapter 4 shows also that non-cognitive abilities may be the cause of the socioeconomic bias observed in school systems because the later are ability-biased. Moreover, the creative potential studied in Chapter 2 shows that creative individuals perform differently some tasks at work and at school, importantly our sample of creative pupils have higher grades at school. In addition, chapters 3 and 4 show that school systems have a great influence over motivation and effort. Taken all these evidences together, it seems crucial to address the development of these noncognitive abilities at school - in addition to cognitive abilities. Help students to have better estimates of their self-confidence, favor the development of creative potential, and stimulate sustainable motivation and effort should be part of the education that pupils receive in schools. Promoting thus better decisions, better outcomes and a more equitable society.

Bibliography

- Adams, Joe K.** 1957. "A confidence scale defined in terms of expected percentages." *The American journal of psychology*, 432–436. (p. [7](#), [41](#), [132](#), and [262](#))
- Ai, Chunrong, and Edward C Norton.** 2003. "Interaction terms in logit and probit models." *Economics letters*, 80(1): 123–129. (p. [87](#))
- Ames, Carole, and Jennifer Archer.** 1988. "Achievement goals in the classroom: Students' learning strategies and motivation processes." *Journal of educational psychology*, 80(3): 260. (p. [17](#), [145](#), and [273](#))
- Anderson, Cameron, Sebastien Brion, Don A Moore, and Jessica A Kennedy.** 2012. "A status-enhancement account of overconfidence." *Journal of personality and social psychology*, 103(4): 718. (p. [68](#))
- Arenas, Andreu, and Clément Malgouyresy.** 2017. "Countercyclical School Attainment and Intergenerational Mobility." *Mimeo, Université catholique de Louvain*. (p. [210](#) and [225](#))
- Armantier, Olivier, and Nicolas Treich.** 2013. "Eliciting beliefs: Proper scoring rules, incentives, stakes and hedging." *European Economic Review*, 62: 17–40. (p. [41](#))
- Armantier, Olivier, Louis Lévy-Garboua, Claire Owen, and Lætitia Placido.** 2015. "Discovering Preferences: A Theoretical Framework and an Experiment." (p. [75](#))
- Arrow, Kenneth J.** 1973. "Higher education as a filter." *Journal of public economics*, 2(3): 193–216. (p. [144](#))
- Arum, Richard, and Yossi Shavit.** 1995. "Secondary vocational education and the transition from school to work." *Sociology of Education*, 187–204. (p. [127](#))
- Bandura, Albert.** 1986. "Social foundations of thought and action: A social cognitive theory." *NY.: Prentice-Hall*. (p. [8](#) and [263](#))

- Bandura, Albert.** 1993. "Perceived self-efficacy in cognitive development and functioning." *Educational psychologist*, 28(2): 117–148. (p. 6 and 261)
- Barber, Brad M, and Terrance Odean.** 2001. "Boys will be boys: Gender, overconfidence, and common stock investment." *Quarterly journal of Economics*, 261–292. (p. 68)
- Barbot, Baptiste, Maud Besançon, and Todd Lubart.** 2011. "Assessing creativity in the classroom." *The Open Education Journal*, 4(1). (p. 9, 74, 247, and 264)
- Baudelot, Christian, and Roger Establet.** 1992. "Allez les filles." *Paris, Seuil*. (p. 186)
- Becker, Gary S.** 1964. *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago press. (p. 19, 144, 186, and 276)
- Becker, Gary S, and Nigel Tomes.** 1979. "An equilibrium theory of the distribution of income and intergenerational mobility." *The Journal of Political Economy*, 1153–1189. (p. 4, 20, 186, 187, 258, and 276)
- Becker, Gary Stanley.** 1967. *Human capital and the personal distribution of income: An analytical approach*. Institute of Public Administration. (p. 19, 186, and 276)
- Bénabou, Roland, and Jean Tirole.** 2002. "Self-confidence and personal motivation." *The Quarterly Journal of Economics*, 117(3): 871–915. (p. 6 and 261)
- Benoît, Jean-Pierre, and Juan Dubra.** 2011. "Apparent overconfidence." *Econometrica*, 79(5): 1591–1625. (p. 32)
- Benoît, Jean-Pierre, Juan Dubra, and Don A Moore.** 2015. "Does the better-than-average effect show that people are overconfident?: Two experiments." *Journal of the European Economic Association*, 13(2): 293–329. (p. 49)
- Berlin, Noémi, Jean-Louis Tavani, and Maud Besançon.** 2016. "An exploratory study of creativity, personality and schooling achievement." *Education Economics*, 24(5): 536–556. (p. 82 and 83)

- Besançon, Maud, and Todd Lubart.** 2008. "Differences in the development of creative competencies in children schooled in diverse learning environments." *Learning and individual differences*, 18(4): 381–389. (p. 10 and 265)
- Black, Sandra E., and Paul J. Devereux.** 2011. "Chapter 16 - Recent Developments in Intergenerational Mobility*." In . Vol. 4, Part B of *Handbook of Labor Economics*, , ed. David Card and Orley Ashenfelter, 1487 – 1541. Elsevier. (p. 19 and 275)
- Blair, Clancy, and Rachel Peters Razza.** 2007. "Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten." *Child development*, 78(2): 647–663. (p. 5 and 260)
- Bonin, Holger, Thomas Dohmen, Armin Falk, David Huffman, and Uwe Sunde.** 2007. "Cross-sectional earnings risk and occupational sorting: The role of risk attitudes." *Labour Economics*, 14(6): 926–937. (p. 79)
- Borghans, Lex, Angela Lee Duckworth, James J Heckman, and Bas Ter Weel.** 2008. "The economics and psychology of personality traits." *Journal of human Resources*, 43(4): 972–1059. (p. 3, 9, 72, 73, 144, 257, and 264)
- Borman, Geoffrey, and Maritza Dowling.** 2010. "Schools and inequality: A multilevel analysis of Coleman's equality of educational opportunity data." *Teachers College Record*, 112(5): 1201–1246. (p. 20 and 277)
- Borst, Grégoire, Amandine Dubois, and Todd I Lubart.** 2006. "Structures et mécanismes cérébraux sous tendant la créativité: une revue de la littérature." *Approche neuropsychologique des apprentissages de l'enfant (ANAE)*, 18(87): 96–113. (p. 10 and 265)
- Boudon, Raymond.** 1973. "L'inégalité des chances: la mobilité sociale dans les sociétés industrielles." (p. 20, 186, and 276)
- Boudon, Raymond.** 1977. "La logique de la frustration relative." *European Journal of Sociology*, 18(01): 3–26. (p. 154)
- Bourdieu, Pierre, and Jean-Claude Passeron.** 1964. *Les héritiers: les étudiants et la culture*. Minuit. (p. 4 and 258)

- Bourdieu, Pierre, and Jean-Claude Passeron.** 1984. "Les héritiers: les étudiants et la culture." (p. 186)
- Bowles, Samuel, Herbert Gintis, and Melissa Osborne.** 2001. "The determinants of earnings: A behavioral approach." *Journal of economic literature*, 39(4): 1137–1176. (p. 1, 3, 73, 255, 257, and 258)
- Breen, Richard.** 2001. "A rational choice model of educational inequality." *Estudios/Working Papers (Centro de Estudios Avanzados en Ciencias Sociales)*, , (166): 1. (p. 68)
- Breen, Richard, and John H Goldthorpe.** 2001. "Class, mobility and merit the experience of two British birth cohorts." *European sociological review*, 17(2): 81–101. (p. 186)
- Brookhart, Susan M, Janet M Walsh, and Wayne A Zientarski.** 2006. "The dynamics of motivation and effort for classroom assessments in middle school science and social studies." *Applied Measurement in Education*, 19(2): 151–184. (p. 13 and 269)
- Brunnermeier, Markus K., and Jonathan A. Parker.** 2005. "Optimal Expectations." *American Economic Review*, 95(4): 1092–1118. (p. 35)
- Buehler, Roger, Dale Griffin, and Michael Ross.** 2002. "Inside the planning fallacy: The causes and consequences of optimistic time predictions." *Heuristics and biases: The psychology of intuitive judgment*, 250–270. (p. 36)
- Busato, Vittorio V, Frans J Prins, Jan J Elshout, and Christiaan Hamaker.** 2000. "Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education." *Personality and Individual differences*, 29(6): 1057–1068. (p. 17, 144, and 273)
- Camerer, Colin.** 2015. "The promise and success of lab-field generalizability in experimental economics: A critical reply to Levitt and List." In *Handbook of Experimental Economic Methodology*. , ed. Guillaume Fréchette and Andrew Schotter, Chapter 26, 249–295. Oxford:Oxford University Press. (p. 22, 174, 210, 224, and 278)

- Camerer, Colin, and Dan Lovallo.** 1999. "Overconfidence and excess entry: An experimental approach." *American economic review*, 306–318. (p. 68)
- Cameron, Stephen V, and Christopher Taber.** 2004. "Estimation of educational borrowing constraints using returns to schooling." *Journal of political Economy*, 112(1): 132–182. (p. 20, 186, and 276)
- Carneiro, Pedro, and James J Heckman.** 2002. "The evidence on credit constraints in post-secondary schooling." *The Economic Journal*, 112(482): 705–734. (p. 19, 186, and 276)
- Castillo, Marco, Paul J Ferraro, Jeffrey L Jordan, and Ragan Petrie.** 2011. "The today and tomorrow of kids: Time preferences and educational outcomes of children." *Journal of Public Economics*, 95(11): 1377–1385. (p. 16, 145, and 273)
- Charness, Gary, and Daniela Grieco.** 2013. "Individual creativity, ex-ante goals and financial incentives." *mimeo*. (p. 8, 11, 12, 263, 266, and 268)
- Chudgar, Amita, and Thomas F Luschei.** 2009. "National income, income inequality, and the importance of schools: A hierarchical cross-national comparison." *American Educational Research Journal*, 46(3): 626–658. (p. 20 and 277)
- Clark, Jeremy, and Lana Friesen.** 2009. "Overconfidence in Forecasts of Own Performance: An Experimental Study*." *The Economic Journal*, 119(534): 229–251. (p. 7 and 42)
- Cole, James S, Louis Rocconi, and Robert Gonyea.** 2012. "Accuracy of self-reported grades: Implications for research." (p. 100)
- Costa, Paul T, and Robert R McCrae.** 1992. *Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO FFI): Professional manual*. Psychological Assessment Resources. (p. 73)
- Costa, Paul T, Robert R McCrae, and John L Holland.** 1984. "Personality and vocational interests in an adult sample." *Journal of Applied psychology*, 69(3): 390. (p. 78)

- Cunha, Flavio, and James J Heckman.** 2008. "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation." *Journal of human resources*, 43(4): 738–782. (p. [20](#), [187](#), and [276](#))
- Cunha, Flavio, James J Heckman, and Susanne M Schennach.** 2010. "Estimating the technology of cognitive and noncognitive skill formation." *Econometrica*, 78(3): 883–931. (p. [4](#), [20](#), [187](#), [258](#), and [276](#))
- Deaves, Richard, Erik Lüders, and Michael Schröder.** 2010. "The dynamics of overconfidence: Evidence from stock market forecasters." *Journal of Economic Behavior & Organization*, 75(3): 402–412. (p. [35](#))
- Deci, Edward L., and Richard M. Ryan.** 1985. "Conceptualizations of Intrinsic Motivation and Self-Determination." *Intrinsic Motivation and Self-Determination in Human Behavior*, 11–40. Boston, MA:Springer US. (p. [12](#), [267](#), and [268](#))
- Deci, Edward L, and Richard M Ryan.** 2010. *Self-determination*. Wiley Online Library. (p. [12](#), [267](#), and [268](#))
- DeGroot, Morris.** 1970. *Optimal Statistical Decisions*. McGraw-Hill. (p. [58](#))
- Denison, Edward Fulton, et al.** 1962. "Sources of economic growth in the United States and the alternatives before us." (p. [73](#))
- Diamond, Adele.** 2013. "Executive functions." *Annual review of psychology*, 64: 135–168. (p. [2](#), [4](#), [5](#), [256](#), [258](#), and [260](#))
- Diamond, Adele, and Kathleen Lee.** 2011. "Interventions shown to aid executive function development in children 4 to 12 years old." *Science*, 333(6045): 959–964. (p. [2](#), [4](#), [256](#), and [259](#))
- Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G Wagner.** 2011. "Individual risk attitudes: Measurement, determinants, and behavioral consequences." *Journal of the European Economic Association*, 9(3): 522–550. (p. [78](#))

- Duckworth, Angela L., and Martin E.P. Seligman.** 2005. "Self-Discipline Outdoes IQ in Predicting Academic Performance of Adolescents." *Psychological Science*, 16(12): 939–944. PMID: 16313657. (p. 3 and 257)
- Edwards, Richard C.** 1977. "Personal traits and " success" in schooling and work." *Educational and Psychological Measurement*, 37(1): 125–138. (p. 76)
- Erev, Ido, Thomas S Wallsten, and David V Budescu.** 1994. "Simultaneous over-and underconfidence: The role of error in judgment processes." *Psychological review*, 101(3): 519. (p. 34 and 52)
- Farah, Martha J, Kimberly G Noble, and Hallam Hurt.** 2006. "Poverty, privilege, and brain development: Empirical findings and ethical implications." *Neuroethics: Defining the issues in theory, practice, and policy*, 277–288. (p. 4 and 259)
- Feinstein, JS.** 2009. "Creativity & The Economic System." *Encyclopedia of Giftedness, Creativity, and Talent. New York: Sage Publications.* (p. 8 and 263)
- Filippin, Antonio, and Marco Paccagnella.** 2012. "Family background, self-confidence and economic outcomes." *Economics of Education Review*, 31(5): 824–834. (p. 7, 17, 18, 174, 225, 251, 252, 262, and 274)
- Fredriksson, Peter.** 1997. "Economic incentives and the demand for higher education." *The Scandinavian Journal of Economics*, 99(1): 129–142. (p. 144)
- Fuchs, Thomas, and Ludger Woessmann.** 2004. "What accounts for international differences in student performance?" CESifo working paper. (p. 20 and 276)
- Gagné, Marylène, and Edward L Deci.** 2005. "Self-determination theory and work motivation." *Journal of Organizational behavior*, 26(4): 331–362. (p. 12, 13, 267, and 268)
- Gamoran, Adam.** 1987. "The stratification of high school learning opportunities." *Sociology of education*, 135–155. (p. 127)
- García, Emma.** 2013. "What we learn in school: Cognitive and non-cognitive skills in the educational production function." PhD diss. Columbia University. (p. 73)

- Garcia, Emma.** 2014. "The Need to Address Noncognitive Skills in the Education Policy Agenda. Briefing Paper# 386." *Economic Policy Institute*. (p. 73)
- García, Emma.** 2016. "The Need to Address Non-Cognitive Skills in the Education Policy Agenda." In *Non-cognitive Skills and Factors in Educational Attainment*. 31–64. Springer. (p. 3, 4, 5, 258, and 260)
- Gervais, Simon, and Terrance Odean.** 2001. "Learning to be overconfident." *Review of Financial studies*, 14(1): 1–27. (p. 34)
- Glover, John A, and Fred Sautter.** 1977. "Relation of four components of creativity to risk-taking preferences." *Psychological reports*, 41(1): 227–230. (p. 84)
- Gneezy, Uri, Muriel Niederle, and Aldo Rustichini.** 2003. "Performance in competitive environments: Gender differences." *The Quarterly Journal of Economics*, 118(3): 1049–1074. (p. 200)
- Goodie, Adam S.** 2005. "The role of perceived control and overconfidence in pathological gambling." *Journal of Gambling Studies*, 21(4): 481–502. (p. 68)
- Gosling, Samuel D, Peter J Rentfrow, and William B Swann.** 2003. "A very brief measure of the Big-Five personality domains." *Journal of Research in personality*, 37(6): 504–528. (p. 78)
- Goux, Dominique, Marc Gurgand, and Eric Maurin.** 2016. "Adjusting your dreams? High school plans and dropout behaviour." *The Economic Journal*. (p. 18, 174, 225, and 274)
- Grieco, Daniela, and Robin M Hogarth.** 2009. "Overconfidence in absolute and relative performance: The regression hypothesis and Bayesian updating." *Journal of Economic Psychology*, 30(5): 756–771. (p. 32)
- Griffin, Dale, and Amos Tversky.** 1992. "The weighing of evidence and the determinants of confidence." *Cognitive psychology*, 24(3): 411–435. (p. 32)
- Guilford, JP.** 1950. "Creativity." *American Psychologist*, 5(9): 444–454. (p. 9 and 264)

- Hanushek, Eric A, and Ludger Wößmann.** 2006. “Does early tracking affect educational inequality and performance? Differences-in-differences evidence across countries.” *Economic Journal*, 116(510): C63–C76. (p. [127](#), [186](#), [210](#), and [225](#))
- Hart, Betty, and Todd R Risley.** 1995. *Meaningful differences in the everyday experience of young American children*. Paul H Brookes Publishing. (p. [4](#) and [258](#))
- Hausman, Angela, and Wesley J Johnston.** 2014. “The role of innovation in driving the economy: Lessons from the global financial crisis.” *Journal of Business Research*, 67(1): 2720–2726. (p. [75](#))
- Heath, Chip, and Amos Tversky.** 1991. “Preference and belief: Ambiguity and competence in choice under uncertainty.” *Journal of risk and uncertainty*, 4(1): 5–28. (p. [57](#))
- Heckman, Daniel A, and Martha J Farah.** 2009. “Socioeconomic status and the developing brain.” *Trends in cognitive sciences*, 13(2): 65–73. (p. [20](#), [187](#), and [276](#))
- Heckman, James J.** 2011. “The Economics of Inequality: The Value of Early Childhood Education.” *American Educator*, 35(1): 31. (p. [3](#), [20](#), [258](#), and [276](#))
- Heckman, James J, and Tim Kautz.** 2012. “Hard evidence on soft skills.” *Labour economics*, 19(4): 451–464. (p. [3](#) and [257](#))
- Heckman, James J, and Yona Rubinstein.** 2001. “The importance of noncognitive skills: Lessons from the GED testing program.” *The American Economic Review*, 91(2): 145–149. (p. [2](#), [16](#), [145](#), [255](#), and [273](#))
- Heckman, James J, John Eric Humphries, and Gregory Veramendi.** 2016. “Returns to education: The causal effects of education on earnings, health and smoking.” National Bureau of Economic Research. (p. [15](#), [16](#), [271](#), and [272](#))
- Heckman, James J, Jora Stixrud, and Sergio Urzua.** 2006. “The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior.” *Journal of Labor economics*, 24(3): 411–482. (p. [4](#) and [258](#))

- Heckman, James, Rodrigo Pinto, and Peter Savelyev.** 2013. "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes." *The American economic review*, 103(6): 2052–2086. (p. 4 and 259)
- Hennessey, Beth A., and Teresa M. Amabile.** 2010. "Creativity." *Annual Review of Psychology*, 61(1): 569–598. PMID: 19575609. (p. 12 and 268)
- Holland, John L.** 1966. *The psychology of vocational choice*. (p. 78 and 223)
- Hollard, Guillaume, Sébastien Massoni, and Jean-Christophe Vergnaud.** 2015. "In search of good probability assessors: an experimental comparison of elicitation rules for confidence judgments." *Theory and Decision*, 1–25. (p. 7 and 42)
- Humphries, John Eric, and Fabian Kosse.** 2017. "On the interpretation of non-cognitive skills—what is being measured and why it matters." *Journal of Economic Behavior & Organization*, 136: 174–185. (p. 2, 21, 256, and 277)
- Jaeggi, Susanne M, Martin Buschkuhl, John Jonides, and Walter J Perrig.** 2008. "Improving fluid intelligence with training on working memory." *Proceedings of the National Academy of Sciences*, 105(19): 6829–6833. (p. 2 and 256)
- Jencks, Christopher.** 1973. "Inequality." (p. 186)
- Johnson, Dominic D. P.** 2004. *Overconfidence and War: The Havoc and Glory of Positive Illusions*. Cambridge, MA: Harvard UP. (p. 68)
- Johnson, Dominic DP, and James H Fowler.** 2011. "The evolution of overconfidence." *Nature*, 477(7364): 317–320. (p. 68)
- Kahneman, Daniel, and Amos Tversky.** 1979. "Prospect theory: An analysis of decision under risk." *Econometrica: Journal of the econometric society*, 263–291. (p. 34)
- Karwowski, Maciej, Jan Dul, Jacek Gralewski, Emanuel Jauk, Dorota M Jankowska, Aleksandra Gajda, Michael H Chruszczewski, and Mathias Benedek.** 2016. "Is creativity without intelligence possible? A Necessary Condition Analysis." *Intelligence*, 57: 105–117. (p. 75)

- Köszegi, Botond.** 2006. "Ego utility, overconfidence, and task choice." *Journal of the European Economic Association*, 4(4): 673–707. (p. 35)
- Król, Anna, Marta Dziechciarz-Duda, et al.** 2013. "On the non-monetary benefits of tertiary education." *Ekonometria*, , (41): 78–94. (p. 16 and 272)
- Kruger, Justin.** 1999. "Lake Wobegon be gone! The " below-average effect" and the egocentric nature of comparative ability judgments." *Journal of personality and social psychology*, 77(2): 221–232. (p. 6, 32, 197, and 261)
- Kruger, Justin, and David Dunning.** 1999. "Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments." *Journal of personality and social psychology*, 77(6): 1121–1134. (p. 35 and 46)
- Lance, Lochner.** 2011. "Chapter 2 - Nonproduction Benefits of Education: Crime, Health, and Good Citizenship." In *Handbook of The Economics of Education*. Vol. 4 of *Handbook of the Economics of Education*, , ed. Stephen Machin Eric A. Hanushek and Ludger Woessmann, 183 – 282. Elsevier. (p. 16 and 272)
- Langer, Ellen J, and Jane Roth.** 1975. "Heads I win tails it's chance: The illusion of control as a function of the sequence outcomes in a purely chance task." *Journal of Personality and Social Psychology*, 32: 951–955. (p. 34)
- Lassibille, Gérard, and Lucía Navarro-Gómez.** 2000. "Organisation and Efficiency of Education Systems: Some Empirical Findings." *Comparative Education*, 36(1): 7–19. (p. 126)
- Levitt, Steven D, and John A List.** 2007. "What do laboratory experiments measuring social preferences reveal about the real world?" *The journal of economic perspectives*, 21(2): 153–174. (p. 21, 22, and 278)
- Lévy-Garboua, Louis, Nathalie Damoiselet, Gérard Lassibille, and Lucia Navarro-Gomez.** 2004. "An economist's view of schooling systems." *Human capital over the life cycle*, 53–68. (p. 126)
- Lichtenstein, Sarah, and Baruch Fischhoff.** 1977. "Do those who know more also know more about how much they know?" *Organizational behavior and human performance*, 20(2): 159–183. (p. 6, 32, 159, and 261)

- Lichtenstein, Sarah, and Paul Slovic.** 1971. "Reversals of preference between bids and choices in gambling decisions." *Journal of experimental psychology*, 89(1): 46–55. (p. 68)
- Lichtenstein, S, B Fischhoff, and L Phillips.** 1982. "Calibration of probabilities: The state of the art to 1980. D. Kahneman, P. Slovic, and A. Tverski (Eds.) *Judgment under uncertainty: Heuristics and biases.*" (p. 6, 32, and 261)
- Lubart, TI, Franck Zenasni, and Baptiste Barbot.** 2013. "Creative potential and its measurement." *International Journal of Talent Development and Creativity*, 1(2): 41–51. (p. 9, 23, 264, and 280)
- Lubart, Todd, and Jacques-Henri Guignard.** 2004. "The Generality-Specificity of Creativity: A Multivariate Approach." (p. 9 and 264)
- Lubart, Todd, C. Mouchiroud, S. Tordjman, and F. et Zenasni.** 2003. *Psychologie de la créativité.* Armand Colin. (p. 9 and 265)
- Lubart, Todd I, and Robert J Sternberg.** 1995. "An investment approach to creativity: Theory and data." *The creative cognition approach*, 269–302. (p. 10 and 265)
- Lubart, Todd, Maud Besançon, and Baptiste Barbot.** 2011. "Evaluation du potentiel créatif (EPoC)." *Paris, France: Editions Hogrefe.* (p. 11, 79, 82, 84, 222, and 266)
- Lévy-Garboua, Louis.** 1973. "Rémunère-t-on les études?" *Consommation*, 3: 57–75. (p. 20, 187, and 276)
- Malmendier, Ulrike, and Geoffrey Tate.** 2005. "CEO overconfidence and corporate investment." *The journal of finance*, 60(6): 2661–2700. (p. 68)
- McCrae, Robert R, and Paul T Costa.** 1987. "Validation of the five-factor model of personality across instruments and observers." *Journal of personality and social psychology*, 52(1): 81. (p. 11, 79, 84, and 267)
- Menger, Pierre-Michel, and Steven Rendall.** 2014. *The Economics of Creativity.* Harvard University Press. (p. 8, 72, and 263)

- Merkle, Christoph, and Martin Weber.** 2011. "True overconfidence: The inability of rational information processing to account for apparent overconfidence." *Organizational Behavior and Human Decision Processes*, 116(2): 262–271. (p. 49)
- Miller, Dale T, and Michael Ross.** 1975. "Self-serving biases in the attribution of causality: Fact or fiction?" *Psychological bulletin*, 82(2): 213. (p. 34)
- Miller, Tyler M, and Lisa Geraci.** 2011. "Unskilled but aware: reinterpreting overconfidence in low-performing students." *Journal of experimental psychology: learning, memory, and cognition*, 37(2): 502. (p. 47)
- Mincer, Jacob.** 1974. "Schooling, Experience, and Earnings. Human Behavior & Social Institutions No. 2." (p. 73)
- Möbius, Markus M, Muriel Niederle, Paul Niehaus, and Tanya S Rosenblat.** 2014. "Managing Self-Confidence." Working Paper. (p. 6, 35, and 260)
- Moore, Don A, and Paul J Healy.** 2008. "The trouble with overconfidence." *Psychological review*, 115(2): 502. (p. 6, 32, 34, 52, 57, and 261)
- national Research Council, et al.** 2013. *Education for life and work: Developing transferable knowledge and skills in the 21st century*. National Academies Press. (p. 10 and 265)
- Niederle, Muriel, and Lise Vesterlund.** 2008. "Gender differences in competition." *Negotiation Journal*, 24(4): 447–463. (p. 200)
- Oskamp, Stuart.** 1965. "Overconfidence in case-study judgments." *Journal of consulting psychology*, 29(3): 261. (p. 6, 32, and 261)
- Page, Lionel, Louis Lévy-Garboua, and Claude Montmarquette.** 2007. "Aspiration levels and educational choices: An experimental study." *Economics of Education Review*, 26(6): 747–757. (p. 128 and 201)
- Piketty, Thomas.** 2000. "Theories of persistent inequality and intergenerational mobility." *Handbook of income distribution*, 1: 429–476. (p. 19 and 275)

- Plott, Charles R.** 1991. "Will economics become an experimental science?" *Southern Economic Journal*, 901–919. (p. 22 and 279)
- Powell, Lynda Elaine, and Anita M Myers.** 1995. "The activities-specific balance confidence (ABC) scale." *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 50(1): M28–M34. (p. 8 and 263)
- Raiffa, Howard.** 1968. *Decision Analysis: Introductory Lectures on Choices und Uncertainty*. Addison-Wesley. (p. 15, 145, and 270)
- Roese, Neal J, and Amy Summerville.** 2005. "What we regret most... and why." *Personality and Social Psychology Bulletin*, 31(9): 1273–1285. (p. 16, 146, and 272)
- Romer, Paul M.** 1986. "Increasing returns and long-run growth." *The journal of political economy*, 1002–1037. (p. 8, 75, and 263)
- Ruhose, Jens, and Guido Schwerdt.** 2016. "Does early educational tracking increase migrant-native achievement gaps? Differences-in-differences evidence across countries." *Economics of Education Review*, 52: 134–154. (p. 186)
- Ryan, Richard M, and Edward L Deci.** 2000. "Intrinsic and extrinsic motivations: Classic definitions and new directions." *Contemporary educational psychology*, 25(1): 54–67. (p. 12 and 268)
- Ryvkin, Dmitry, Marian Krajč, and Andreas Ortmann.** 2012. "Are the unskilled doomed to remain unaware?" *Journal of Economic Psychology*, 33(5): 1012–1031. (p. 46)
- Sander, Paul, and Lalage Sanders.** 2003. "Measuring confidence in academic study: A summary report." *Electronic Journal of Research in Educational Psychology and Psychopedagogy*, 1(1): 1–17. (p. 8 and 262)
- Sander, Paul, and Lalage Sanders.** 2006. "Understanding Academic Confidence." *Psychology Teaching Review*, 12(1): 29–42. (p. 8 and 263)
- Schultz, Theodore W.** 1975. "The value of the ability to deal with disequilibria." *Journal of economic literature*, 13(3): 827–846. (p. 76)

- Schumpeter, Joseph A.** 1934. "The Theory of economic development." (p. 76)
- Schwartz, Barry.** 2004. "The paradox of choice: Why more is less." Ecco New York. (p. 15 and 271)
- Schwartz, Barry.** 2009. "Incentives, choice, education and well-being." *Oxford Review of Education*, 35(3): 391–403. (p. 17 and 273)
- Shiller, Robert J.** 2000. "Measuring bubble expectations and investor confidence." *The Journal of Psychology and Financial Markets*, 1(1): 49–60. (p. 68)
- Siegel, Linda S.** 1989. "IQ is irrelevant to the definition of learning disabilities." *Journal of learning disabilities*, 22(8): 469–478. (p. 135)
- Spence, Michael.** 1973. "Job market signaling." *The quarterly journal of Economics*, 355–374. (p. 126 and 144)
- Stankov, Lazar.** 1999. "Mining on the "no man's land" between intelligence and personality." *Learning and individual differences: Process, trait, and content determinants*, 315–337. (p. 6 and 261)
- Stankov, Lazar, Jihyun Lee, Wenshu Luo, and David J Hogan.** 2012. "Confidence: A better predictor of academic achievement than self-efficacy, self-concept and anxiety?" *Learning and Individual Differences*, 22(6): 747–758. (p. 6, 7, 8, 261, and 263)
- Sternberg, Robert J.** 2000. *Practical intelligence in everyday life*. Cambridge University Press. (p. 73)
- Sternberg, Robert J.** 2001. "What is the common thread of creativity? Its dialectical relation to intelligence and wisdom." *American Psychologist*, 56(4): 360. (p. 73)
- Sternberg, Robert J.** 2006. "The nature of creativity." *Creativity research journal*, 18(1): 87–98. (p. 10 and 265)
- Sternberg, Robert J, and Todd I Lubart.** 1991a. "Creating creative minds." *The Phi Delta Kappan*, 72(8): 608–614. (p. 10, 84, and 265)

- Sternberg, Robert J, and Todd I Lubart.** 1991*b*. "An investment theory of creativity and its development." *Human development*, 34(1): 1–31. (p. 10 and 265)
- Sternberg, Robert J, and Todd I Lubart.** 1995. *Defying the crowd: Cultivating creativity in a culture of conformity*. Free Press. (p. 9 and 264)
- Sternberg, Robert J, and Todd I Lubart.** 1996. "Investing in creativity." *American psychologist*, 51(7): 677. (p. 8, 72, and 263)
- Stiglitz, Joseph E.** 1975. "The theory of " screening," education, and the distribution of income." *The American Economic Review*, 65(3): 283–300. (p. 144)
- Stone, Dan N, Edward L Deci, and Richard M Ryan.** 2009. "Beyond talk: Creating autonomous motivation through self-determination theory." *Journal of General Management*, 34(3): 75–91. (p. 13 and 268)
- Storme, Martin, Jean-Louis Tavani, and Nils Myszkowski.** 2016. "Psychometric Properties of the French Ten-Item Personality Inventory (TIPI)." *Journal of Individual Differences*. (p. 78)
- Stouffer, Samuel A, Edward A Suchman, Leland C DeVinney, Shirley A Star, and Robin M Williams Jr.** 1949. "The American soldier: Adjustment during army life.(Studies in social psychology in World War II), Vol. 1." (p. 154)
- Svenson, Ola.** 1981. "Are we all less risky and more skillful than our fellow drivers?" *Acta Psychologica*, 47(2): 143–148. (p. 6, 32, and 261)
- Syed, Jawad, Ronald J Burke, F Pinar Acar, Ayala Malach Pines, Miri Lerner, and Dafna Schwartz.** 2010. "Gender differences in entrepreneurship: equality, diversity and inclusion in times of global crisis." *Equality, Diversity and Inclusion: An International Journal*, 29(2): 186–198. (p. 95)
- Taylor, Jeannette, and Ranald Taylor.** 2011. "Working hard for more money or working hard to make a difference? Efficiency wages, public service motivation, and effort." *Review of Public Personnel Administration*, 31(1): 67–86. (p. 13 and 269)
- Torrance, E Paul.** 1962. "Guiding creative talent." (p. 11, 79, and 267)

- Tversky, Amos, and Daniel Kahneman.** 1973. "Availability: A heuristic for judging frequency and probability." *Cognitive psychology*, 5(2): 207–232. (p. [53](#))
- Vallerand, Robert J, Luc G Pelletier, Marc R Blais, Nathalie M Briere, Caroline Senecal, and Evelyne F Vallieres.** 1992. "The Academic Motivation Scale: A measure of intrinsic, extrinsic, and amotivation in education." *Educational and psychological measurement*, 52(4): 1003–1017. (p. [13](#) and [269](#))
- Van den Steen, Eric.** 2011. "Overconfidence by Bayesian-rational agents." *Management Science*, 57(5): 884–896. (p. [6](#), [32](#), and [260](#))
- Vila, Luis E.** 2000. "The Non-Monetary Benefits of Education." *European Journal of Education*, 35(1): 21–32. (p. [16](#) and [272](#))
- Weiss, Andrew.** 1983. "A sorting-cum-learning model of education." *The Journal of Political Economy*, 420–442. (p. [144](#))
- Wößmann, Ludger.** 2016. "The Importance of School Systems: Evidence from International Differences in Student Achievement." *The Journal of Economic Perspectives*, 30(3): 3–31. (p. [14](#), [127](#), and [270](#))

List of Tables

1.1	Descriptive statistics for the three treatments	39
1.2	Descriptive statistics by ability level	40
1.3	A comparison of confidence for the wall and hill treatments shown separately	44
1.4	OLS estimation of the Bayesian model of confidence before round 5	61
1.5	OLS estimation of the Bayesian model of confidence for doublers reported before the middle level	62
1.6	Estimation of posterior confidence (after doubling) and ex post chances of success at the middle level	65
1.7	The prevalence and cost of miscalibration among doublers	66
2.1	Aspects in the production of creative work (Barbot, Besançon and Lubart, 2011)	74
2.2	Normalized creativity scores	83
2.3	Creativity correlation table with p-values in parentheses (Pearson's test)	84
2.4	Creative agents' characteristics (correlations)	85
2.5	Typing task: outcomes and efficiency	87
2.6	Probability of typing in disorder	89
2.7	Typing task: number of correct codes	90
2.8	Probability of producing more than the partner	91
2.9	Buttons task: number of points (OLS)	92
2.10	Buttons task (no jackpot)	97
2.11	Finding the jackpot	98
2.12	Satisfaction	98
2.13	Task satisfaction	99
2.14	OLS regression of school achievements of French 9 th graders	102
2.15	Descriptive statistics	105
2.16	Variables description: buttons task ($n = 169$)	105
2.17	Variables description: typing task ($n = 109$)	107
2.18	Other variables description	107
2.19	Number of correct codes in the typing task	107

2.20	Grades: descriptive statistics ($n = 119$)	108
2.21	Correlations between creativity measures and personality variables. . .	109
2.22	Typing task: outcomes and efficiency	110
2.23	Buttons task: vocational scores	111
2.24	Satisfaction (same sample of buttons task).	112
2.25	OLS regression of school achievements of French 9 th graders	113
2.26	Holland's typology	114
II.1	Descriptive statistics by treatment	134
II.2	Statistical tests for Table II.1	134
II.3	Level attained	136
II.4	Outcomes	136
3.1	Efficiency of treatments	148
3.2	Statistical tests for Table 3.1	148
3.3	Chances of success for secondary level	151
3.4	Comparison to No-choice outcomes	152
3.5	Efficiency of treatments	156
3.6	Statistical tests for Table 3.5	157
3.7	Incentives and efficiency of school systems	165
3.8	Global performance	168
3.9	Performance at the compulsory level	169
3.10	Performance at post-compulsory levels	170
3.11	Self-confidence for secondary level	171
3.12	Efficiency of treatments (complete version)	175
3.13	Statistical tests for Table 3.12	175
3.14	Maximum level attained: probability of success	177
3.15	Ex-post chances of success for No-choice participants	179
3.16	Expected utility for different ability levels and risk profiles	180
4.1	Descriptive statistics by ability level	190
4.2	Descriptive statistics by ability level: statistical tests for Table 4.1 . . .	190
4.3	Variation of performance across sorting mechanisms by ability level . .	192
4.4	Expected utility for different ability levels and risk profiles	197
4.5	Decisions according to the risk attitude and ability level	198

4.6	Variation of performance across sorting mechanisms by gender	200
4.7	Variation of the frequency of quits after the primary level, by ability tercile	202
4.8	Variation of outcomes with incentives, by ability tercile	203
4.9	Variation of performance with incentives, by gender	206
4.10	Main outcomes by ability level: statistical tests for Table 4.3	212
4.11	Proportion of participants who decide to stop after succeed the compulsory level.	212
4.12	Proportion of participants who decide to stop after succeed the compulsory level: statistical tests (Table 4.11)	213
4.13	Main outcomes by gender: statistical tests (Table 4.6)	213
4.14	Ex-post chances of success for No-choice participants	217
4.15	Expected utility for different ability levels and risk profiles	218
4.16	Decisions according to the risk attitude	219
4.17	Expected utility for different ability levels and risk profiles (low incentives)	219
4.18	Decisions in the choice-low treatment according to the risk attitude . .	220

List of Figures

.1	High school tracking by ability. Reprinted from Filippin and Paccagnella (2012) with permission.	18
1.1	Decision problem perceived by participants at the start of level 2 of the choice treatment.	37
1.2	Hard-easy effect observed at three levels	45
1.3a	Under-confidence at the training level, by ability.	47
1.3b	Overconfidence at middle level, by ability.	47
1.3c	Overconfidence at high level, by ability.	48
1.4	Variation of confidence with experience, by level of ability: middle level	50
1.5	Variation of confidence with experience, by level of ability: high level .	51
1.6	Number of anagrams solved per round by level of ability	51
1.7	Example of the task screen	69
2.1	Distribution of the three normalized scores	83
2.2	Interaction between DT and CT: additional points in the "best moving strategy" series.	94
2.3	Distribution of scores: DT and IT.	106
2.4	Distribution of correct codes in the typing task	108
2.5	Accumulated distribution of points among buttons.	115
2.6	Order of activities	116
2.7	Examples of divergent thinking graphical test	117
2.8	Examples of convergent thinking graphical test high scores	118
2.9	Examples of convergent thinking graphical test medium scores	118
2.10	Examples of convergent thinking graphical test low scores	119
2.11	Buttons: comprehension questions (in French): "Performance treatment"	120
2.12	Buttons: comprehension questions (in French): "Competition treatment"	120
2.13	Buttons - screen for series 1 and 3	121
2.14	Buttons - screen for series 2	121
II.1	School system: stylized design	127
II.2	Experimental design	130

3.1	Success at secondary level, conditional on starting this level.	149
3.2	Distribution of participants to post-compulsory education.	150
3.3	Choice treatment: number of anagrams solved at secondary level in the GEN	151
3.4	Distribution of participants between the two possible tracks.	166
3.5	Maximum level attained (cumulative distribution)	176
3.6	Maximum level attained: marginal effects (table 3.14)	178
4.1	Ratio of average payoffs and outputs (low/high ability)	194
4.2	Participation of ability groups at each level attained in the task	195
4.3	Distribution of participants between the two possible tracks by ability level.	199
4.4	Ratio of average payoffs and outputs (women/men)	201
4.5	Proportion of genders at each level attained	201
4.6	Distribution of participants between the two possible tracks.	204
4.7	Ratio of average payoffs and outputs (low/high ability)	205
4.8	Participation of the three ability terciles at each level attained in the task.	205
4.9	Ratio of average payoffs and outputs (women/men)	207
4.10	Proportion of participants at each level attained	207
4.11	Low vs. medium-ability	213
4.12	Medium vs. high-ability	214
4.13	Decision tree: probabilities are conditional on prior success.	215
II.1	Suivi du cursus secondaire par aptitude. Réimprimé de <i>Filippin and Paccagnella (2012)</i> avec la permission.	274

Résumé substantiel

Cette thèse contribue à la littérature économique sur les compétences *non cognitives* essentielles pour la réussite de la vie, particulièrement pour la réussite scolaire. Elle comprend quatre essais basés sur des approches économiques comportementales et expérimentales, avec deux objectifs principaux. Le premier objectif est d'étudier deux compétences non cognitives, à savoir la confiance en soi et la créativité. Notre but est alors de comprendre les déterminants de la confiance en soi et l'impact de la créativité sur les résultats économiques. Le deuxième objectif est d'étudier comment le système scolaire influence les décisions éducatives, les résultats scolaires et la mobilité intergénérationnelle, secteurs où les compétences non cognitives peuvent jouer un rôle important, en particulier via leurs effets sur la confiance en soi et la motivation. Ce résumé substantiel met en perspective les questions et concepts développés dans chacun des chapitres composant cette thèse et en donne un aperçu.

1. La définition hasardeuse des compétences non cognitives

Le terme "non cognitif" a émergé dans la littérature économique au début des années 2000 dans le but d'expliquer notamment la variabilité des résultats scolaires, ceux du marché du travail et d'autres résultats économiques inexpliqués par des mesures de compétences cognitives. Les compétences cognitives sont en effet mesurées par des tests d'intelligence, par des notes scolaires ou par des tests standards (voir [Bowles, Gintis and Osborne, 2001](#), [Heckman and Rubinstein, 2001](#), par exemple), qui mesurent

l'intelligence et les connaissances. Cependant, l'identification, la classification et la mesure des compétences non cognitives restent un défi pour les économistes¹ (Humphries and Kosse, 2017). Ainsi, l'apport scientifique de d'autres sciences peut être notoire pour les économistes.

Les neuroscientifiques expliquent que la plupart de ces compétences dépendent des fonctions exécutives du cerveau (Diamond and Lee, 2011). Ce dernier se réfère à une famille de fonctions mentales (contrôle cognitif) qui sont nécessaires lorsque l'agent doit se concentrer, réfléchir au lieu de se référer à "l'automatique", et s'appuie principalement sur le cortex préfrontal. Les trois fonctions exécutives principales sont le contrôle inhibiteur², la mémoire de travail³, et la flexibilité cognitive⁴ (Diamond, 2013). Les fonctions exécutives de haut niveau sont utilisées lors de la résolution des problèmes, du raisonnement et de la planification, qui sont liées à l'intelligence fluide⁵. Ainsi, il semble évident que les compétences, appelées non cognitives par les économistes, dépendent aussi de la cognition⁶.

Ainsi, la distinction entre compétences cognitives et non cognitives nous apparaît inappropriée car elle repose sur une fausse dichotomie. Il n'existe en effet pas de mesures de compétences cognitives qui ne reflètent pas, au moins partiellement, les facteurs non-cognitifs de motivation et d'environnement, alors que, dans le même temps, les mesures des compétences non cognitives dépendent également de facteurs cognitifs et situationnels. Ainsi, parler de compétences non cognitives peut être trompeur. Dès

1. La liste exhaustive des terminologies pour les compétences non cognitives trouvées dans la littérature économique illustre particulièrement la difficulté de les définir. Cette littérature comprend des termes tels que les capacités non cognitives, les compétences douces, les compétences socio-émotionnelles, les habiletés comportementales, le caractère et les traits de personnalité.

2. Qui comprend la maîtrise de soi, la discipline et l'attention sélective.

3. Retenir l'information et la manipuler intellectuellement, ce qui est essentiel pour le raisonnement.

4. Y compris la résolution d'une façon créative des problèmes et la flexibilité cognitive.

5. L'intelligence fluide est la capacité de faire face à de nouveaux problèmes, indépendamment de toute connaissance du passé. Elle est considérée comme l'un des facteurs les plus importants dans l'apprentissage (Jaeggi et al., 2008).

6. Le American Psychological Association Dictionary définit la cognition comme "*toutes les formes de connaissance et de conscience telles que percevoir, concevoir, se rappeler, raisonner, juger, imaginer et résoudre des problèmes*".

lors, bien que cette thèse ne vise pas à apporter une nouvelle terminologie pour ces compétences (non-cognitive), nous gardons la terminologie "*non cognitive*", car elle nous permet alors de nous rapporter à la littérature existante. Nous fournissons cependant une plus définition précise: les compétences non cognitives correspondent à des capacités importantes pour la réussite de la vie, mais différentes du savoir (mesuré par les tests de réussite) et du QI. Toutefois les capacités non cognitives sont affectées par ces dernières. Ces capacités non cognitives varient en fonction de l'environnement et, surtout, peuvent être améliorées au cours de la vie. Nous considérons donc la confiance en soi, la créativité et la motivation des compétences non cognitives. Chacune de ces compétences est discutée dans les sections suivantes plus précisément.

2. Quelles compétences non cognitives sont les plus importantes pour l'éducation, le marché du travail et les réussites de la vie?

Comment réussir sa vie ? Une littérature croissante provenant de la psychologie et de la sociologie apporte des idées à la théorie économique et montre que l'éducation, le marché du travail et les succès dans la vie dépendent de nombreuses compétences, et pas seulement des compétences cognitives mesurées par le QI, les notes scolaires et les tests standards de réussite (Borghans et al., 2008). Heckman and Kautz (2012) montrent que les mesures de compétences cognitives au cours de l'adolescence expliquent moins de 15% du salaire horaire à l'âge de 35 ans. Ils suggèrent que les notes ne sont pas seulement déterminées par des compétences techniques. Par exemple, ils montrent que la discipline, par rapport au QI, compte pour deux fois plus dans la variation des notes finales, même à l'université. (Duckworth and Seligman, 2005). Ainsi, l'étude des compétences non cognitives est devenue un sujet important pour les économistes au cours des quinze dernières années comme l'indique la recherche pionnière de Bowles,

Gintis and Osborne (2001) sur les déterminants des gains - un jalon dans la littérature économique des compétences non cognitives. La vaste liste des compétences non cognitives qui sont importantes pour les résultats économiques comprend par exemple: la confiance en soi, le respect d'autrui, la capacité de construire un consensus, la volonté de tolérer les alternatives, la motivation académique, la confiance académique, la persistance, les capacités de communication, la créativité et le travail d'équipe (Heckman, 2011, García, 2016). Compte tenu de la nouveauté relative du domaine pour les économistes, cette liste est susceptible de croître au fur et à mesure de l'apparition de nouvelles recherches.

Les effets directs des compétences non cognitives sont importants pour tous les aspects de la vie: réussite à l'école et sur le lieu de travail, harmonie de la vie conjugale et à échapper au tabagisme, à la toxicomanie ou à participer à des activités illégales (Heckman, Stixrud and Urzua, 2006). Les effets indirects de ces compétences sont également importants, principalement parce que les compétences non cognitives favorisent le développement cognitif. En d'autres termes, le développement des capacités non cognitives permet en effet d'améliorer les compétences académiques telles que la lecture, l'écriture et la performance mathématique (García, 2016).

Le développement des compétences non cognitives commence dès la petite enfance et a une influence importante sur les caractéristiques familiales et sociétales (Cunha, Heckman and Schennach, 2010, Diamond, 2013, García, 2016). Grâce à la socialisation, les parents éduqués transmettent automatiquement leurs capacités et leurs préférences à leurs enfants (Bourdieu and Passeron, 1964, Becker and Tomes, 1979). Par exemple, L'environnement verbal des parents fournit, aux des enfants dès l'âge de trois ans, prédit fortement la compréhension de la lecture à l'âge de 10 ans (Hart and Risley, 1995). De plus, les enfants des classes sociales supérieures bénéficient d'un environnement favorable à leur développement, ils sont en effet mis en contact plus tardivement avec la violence, la mort, les drogues et le système judiciaire pénale et au contraire bénéficie

dès leur enfance d'un accès facilité à la connaissance des chiffres et des lettres, ainsi qu'à la connaissance de d'autres quartiers de la ville, et peuvent lire les gros titres des journaux contrairement aux enfants des classes défavorisés (Farah, Noble and Hurt, 2006). Cependant, les capacités non cognitives sont inégalement réparties en tant que miroir des inégalités sociales. En supposant des capacités innées également distribuées, les inégalités de compétences non cognitives peuvent expliquer la persistance des inégalités sociales⁷.

Étant donné que les fonctions exécutives - et par conséquent les compétences non cognitives - peuvent être améliorées au cours de la vie, la formation dès l'enfance pourrait constituer un excellent moyen de réduire cette inégalité. En effet, il existe des preuves scientifiques favorisant l'amélioration des fonctions exécutives (et des compétences non cognitives) au cours des premières années scolaires. Par exemple, Heckman, Pinto and Savelyev (2013) montrent l'effet positif du Perry Preschool Program sur le développement des compétences non cognitives des enfants défavorisés⁸. De plus, Diamond and Lee (2011) montrent que les interventions au début de la prise en charge scolaire, comme la formation informatisée, l'exercice physique et les arts martiaux, améliorent le développement des fonctions exécutives avec un avantage notable pour les enfants défavorisés⁹, renforçant ainsi le potentiel de l'éducation infantile dans la lutte contre les inégalités de réussite entre les enfants les plus et les moins favorisés.

La politique actuelle de l'éducation se concentre sur les compétences cognitives. Il reste moins de place pour améliorer les compétences non cognitives, même si elles

7. En supposant que la différence de possibilités peut être neutralisée par les politiques publiques.

8. L'objectif du programme préscolaire de Perry (1962-1967) est de fournir une éducation préscolaire de haute qualité aux enfants afro-américains de classes sociales défavorisées dès l'âge de trois ans. Le programme ne s'est pas traduit par une augmentation des points de QI des participants (augmentation faible pour les femmes, nulle pour les hommes), mais par une amélioration des scores sur les tests de réussite à l'âge de 10 ans. Ce résultat confirme l'importance des compétences non cognitives (et/ou leurs améliorations) dans la réussite scolaire.

9. Les enfants à faible revenu, à mémoire de travail plus faible et les enfants en déficit d'attention et hyperactivité (TDAH) montrent que l'amélioration des fonctions exécutives provient de ces interventions.

peuvent être améliorées dans les écoles (Blair and Razza, 2007), entraînant ainsi une mobilité intergénérationnelle accrue. La littérature sur les compétences non cognitives (et les fonctions exécutives) suggère que l'éducation a de multiples dimensions, englobant les compétences et les attitudes, et pas seulement l'intelligence et la connaissance (Diamond, 2013). Dans l'article "*The Need to Address Non-Cognitive Skills in the Education Policy Agenda.*", García résume ainsi les compétences non cognitives que les écoles devraient développer et que les politiques devraient promouvoir:

"[...] cela inclut les compétences de réflexion critique, les compétences en résolution de problèmes, la santé émotionnelle, les habiletés sociales, l'éthique du travail et la responsabilité de la communauté. Mais aussi de façon toute aussi importante les facteurs qui influencent sur les relations personnelles entre les élèves et les enseignants (proximité, affection et communication ouverte), la maîtrise de soi, l'autorégulation, la persévérance, la confiance académique, le travail d'équipe, les compétences organisationnelles, la créativité et compétences en communication."

Les sous-sections suivantes mettent en perspective l'importance économique, les progrès dans la recherche et la mesure des compétences non cognitives abordées dans cette thèse: confiance en soi, potentiel créatif et motivation.

2.1. La confiance en soi

Les croyances subjectives sont importantes pour toutes les situations où un agent économique prend des décisions en situation d'incertitude. Le décideur attribue des estimations de probabilité subjective pour chaque état de la nature impliqué dans la décision, sous condition de choisir l'état (de la nature) qui maximise son utilité espérée. Lorsque la vraie probabilité est inconnue, l'agent estime que sa probabilité subjective met à jour des informations préalables sur sa capacité et sur l'environnement, comme le font les Bayésiens (Van den Steen, 2011, Möbius et al., 2014). Du point de la théorie

économique standard, la confiance est une probabilité déformée de succès qui peut se réviser en fonction de l'expérience et des informations disponibles. En effet, l'impact de la confiance en soi sur le comportement des agents va au-delà du processus décisionnel, car cela a également une incidence sur la motivation et donc sur l'effort pour accomplir la tâche et en augmente ainsi la probabilité de succès. La confiance en soi constitue alors une incitation à renforcer et à maintenir son estime de soi (Bandura, 1993, Bénabou and Tirole, 2002).

Cette thèse considère un type particulier de confiance: la confiance en soi, c'est-à-dire la croyance qu'un agent détient sur sa propre capacité à réussir. Dans de nombreuses circonstances, les gens semblent être trop confiants dans leurs propres capacités quelle que soit la difficulté de la tâche, c'est-à-dire que leur probabilité subjective de réussite est plus élevée que les chances "normales" de réussir la tâche. Moore and Healy (2008) identifient trois formes différentes de confiance excessive en soi, tel que le surplacement, la surestimation, and la surprécision. Surplacement survient lorsque les individus se comparent avec d'autres, se trouvant «mieux que la moyenne» dans des domaines familiers (p. Ex. Svenson, 1981, Kruger, 1999). La surestimation est la plus fréquente dans la littérature, elle se déroule lorsque les agents surestiment leur propre capacité absolue à effectuer une tâche (p. Ex. Lichtenstein and Fischhoff, 1977, Lichtenstein, Fischhoff and Phillips, 1982). Enfin, la sur-précision survient lorsque les gens surestiment la précision de leurs estimations et prévisions (par exemple Oskamp, 1965). Cette thèse vise à comprendre comment les gens surestiment, ou parfois sous-estiment, leur propre capacité absolue à effectuer une tâche isolément.

L'estimation de la confiance en soi pour effectuer une tâche spécifique dépend de la capacité cognitive et de d'autres caractéristiques individuelles (Stankov, 1999). Ainsi, plus la capacité cognitive est élevée, plus le pourcentage d'estimation de réussite est faible (Stankov et al., 2012). Le genre est un exemple d'une caractéristique individuelle qui affecte la sur-confiance ou la confiance excessive. Stankov et al. (2012) montrent

que les filles présentent un biais d'estimation à la baisse de leur confiance en soi par rapport aux garçons, même si elles obtiennent le même niveau de confiance dans leurs scores de réussite aux tests de mathématiques et d'anglais¹⁰. Une autre caractéristique individuelle qui a une incidence sur la confiance en soi est le contexte familial. En utilisant des données nationales croisées (PISA), [Filippin and Paccagnella \(2012\)](#) montrent la relation positive entre le milieu familial et la confiance académique rapportée par les élèves de 15 ans. Pour un niveau de capacité donné, plus le statut socioéconomique est élevé, plus la confiance subjective en soi est élevée pour réussir à l'école.

Si, d'une part, la surestimation de la confiance en soi peut avoir un impact négatif conduisant les individus à des décisions non optimales, d'autre part, cela accroît aussi l'effort et augmente les performances et la probabilité de réussite. Lorsqu'on associe les effets de la confiance sur les décisions et sur la performance en rapport à l'impact de l'origine familiale sur niveau de confiance en soi, [Filippin and Paccagnella \(2012\)](#) affirment que: «la confiance en soi peut être un canal par lequel les inégalités d'éducation et de gain se perpétuent entre les générations».

Nous étudions la confiance en soi en utilisant une mesure d'auto-évaluation. Les individus sont directement invités à indiquer leur probabilité de succès pour une tâche donnée, par exemple: "*quelles sont vos chances de succès sur l'échelle de 0 à 100?*". L'échelle de Adams ([1957](#)) est pratique pour l'analyse quantitative car elle convertit la confiance en probabilités subjectives (presque) continues. Les méthodes d'auto-évaluation ont été largement utilisées et validées par des psychologues et des neuroscientifiques.

Les psychologues ont développé plusieurs échelles pour obtenir des mesures de confiance en soi dans des domaines spécifiques, qui ne sont pas utilisés dans cette thèse. Par exemple, l'échelle Academic Behavioural Confidence ([Sander and Sanders, 2003](#),

10. en d'autres termes, les garçons réussissent moins que les filles tandis que leur niveau de confiance en soi est plus élevé que celui des filles.

2006) fournit une mesure de la confiance académique et de l'auto-efficacité¹¹. L'auto-efficacité est considérée comme une bonne approximation de la confiance en soi¹², trouvée couramment dans la littérature sur la confiance (voir Stankov et al., 2012, par exemple).

2.2. Le potentiel créatif

La créativité a été définie comme "*la capacité de produire des travaux à la fois novateurs et appropriés*" (Sternberg and Lubart, 1996), ce qui constitue un moteur important de l'innovation. Selon Feinstein (2009) la créativité et sa contrepartie l'innovation sont la racine du progrès et donc sont fondamentales pour la dynamique des systèmes économiques. En effet, plusieurs théories attribuent l'innovation technologique à la forte croissance économique après la Seconde Guerre mondiale (Romer, 1986, par exemple). Ainsi, le potentiel de la créativité, c'est-à-dire le potentiel de produire des œuvres créatives - devrait être un sujet d'intérêt majeur pour les économistes. Cependant, peu d'études économiques ont traité les comportements créatifs jusqu'à présents. La littérature économique existante réside dans le fait que la production et la consommation de nouveaux produits sont des activités incertaines, ce qui implique une prise de risques et des compétences entrepreneuriales associées au comportement créatif (Menger and Rendall, 2014). Plus récemment, Charness and Grieco (2013) a étudié l'effet des incitations sur la production d'œuvres créatives. Cependant l'impact du potentiel de la créativité sur les résultats économiques est encore un manque dans la recherche économique¹³.

11. Un autre exemple: Activities-specific Balance Confidence Scale (Powell and Myers, 1995) utilisée dans le domaine médical mesure la confiance dans l'exécution de diverses activités ambulatoires sans tomber ou éprouver un sentiment d'instabilité.

12. L'auto-efficacité a été définie par Bandura (1986, page 391) comme «les jugements des gens sur leurs capacités à s'organiser et à exécuter les actions nécessaires pour atteindre les performances désirées».

13. Cet impact du potentiel de la créativité sur les résultats économiques est abordé dans cette thèse au chapitre 2.

La rareté de la recherche sur le comportement créatif dans le domaine de l'économie peut être attribuée à deux facteurs principalement. Tout d'abord, les économistes ont négligé l'impact des capacités non cognitives, à savoir la créativité des résultats économiques jusqu'à la dernière décennie (Borghans et al., 2008). Deuxièmement, l'évaluation du potentiel de créativité est toujours un goulet d'étranglement dans la recherche économique, ces mesures ne sont pas encore applicables à grande échelle, étant limitée aux études expérimentales. L'intérêt récent pour les capacités non cognitives et l'évolution de la recherche sur la créativité par les psychologues au cours des trois dernières décennies¹⁴ ont cependant mis la créativité sous les projecteurs de la recherche en économie

Le travail de Guilford (1950) a été un tournant dans la recherche psychologique sur la créativité. Jusqu'à ce travail séminal, la créativité était associée à un processus exceptionnel d'individus doués. Ainsi, l'évaluation de la créativité n'était pas un problème important puisque la «créativité» était directement observable par la production d'œuvres artistiques (Barbot, Besançon and Lubart, 2011). Dans le contexte de la période postérieure à la Seconde Guerre mondiale - qui nécessitait une innovation dans la recherche et le développement -, Guilford (1950) a affirmé que "le potentiel de la créativité n'était pas limité aux individus doués et, surtout, pouvait être mesuré et développé. La créativité est donc considérée comme un processus cognitif et social, et non seulement un trait de personnalité. En effet, le potentiel créatif dépend aussi de la nature des domaines¹⁵ et de la tâche en question (Lubart and Guignard, 2004).

Les théories développées par les psychologues au cours des dernières décennies ont confirmé la proposition de Guilford: un comportement créatif dépend de nombreux facteurs. Sternberg and Lubart (1995) proposent une approche multivariée, pour laquelle

14. Barbot, Besançon and Lubart (2011) affirme que, dans les années 90, "la littérature de recherche sur la créativité a augmenté exponentiellement avec l'apparition de nouvelles revues scientifiques, de conférences internationales et de séries de livres sur le sujet, ce qui a coïncidé d'autre part, avec des progrès significatifs dans la science psychométrique".

15. Exemples de domaines créatifs: graphique-artistique, verbal-literacy (compétence littéraire), résolution de problèmes sociaux, musical et créatif (Lubart, Zenasni and Barbot, 2013).

la créativité est influencée par le cognitif (intelligence et connaissance), ou le conatif (motivation, traits de personnalité et style de pensée) et les facteurs environnementaux¹⁶. Dans le même ordre d'idées, la théorie de l'investissement de la créativité énumère six ressources distinctes mais interdépendantes nécessaires pour renforcer la créativité, à savoir les capacités intellectuelles, les connaissances¹⁷, les styles de pensée, la personnalité, la motivation et l'environnement (Sternberg and Lubart, 1991b, Lubart and Sternberg, 1995, Sternberg, 2006). La perspective neurologique attribue au cortex frontal et préfrontal, donc aux fonctions exécutives, un rôle central dans le processus créatif (Borst, Dubois and Lubart, 2006).

Ces théories montrent principalement que le potentiel de la créativité peut être amélioré. Par conséquent, les écoles ont une influence importante sur le développement du comportement créatif. L'environnement d'apprentissage et la pédagogie ont un impact direct sur le développement du potentiel créatif, ce potentiel étant plus élevé pour les enfants ayant un potentiel créatif initial inférieur (Besançon and Lubart, 2008). Par exemple, Sternberg and Lubart (1991a) montrent que les pédagogies alternatives, comme Montessori et Freinet, peuvent développer la pensée critique, une composante importante de la créativité. Dans ce contexte, les auteurs affirment que *"la scolarité peut créer des esprits créatifs - bien que ce ne soit souvent pas le cas"*. Ainsi, le développement d'un environnement d'apprentissage pour améliorer la créativité, considérée par le National Research Council (2013) comme l'une des principales compétences nécessaires pour les résultats d'apprentissage du 21ème siècle, semble être un objectif important de l'éducation.

En psychologie, le processus décisionnel créatif se décompose en une phase de divergence mentale suivie d'une phase de convergence mentale. La divergence mentale

16. Des facteurs émotionnels peuvent être ajoutés à cette liste (Lubart et al., 2003)

17. Les connaissances peuvent à la fois promouvoir ou inhiber la créativité. D'une part, il est impossible d'innover dans un domaine si on ne connaît pas cela, et d'autre part, si l'on connaît beaucoup de connaissances sur un domaine il peut en résulter une perspective fermée (Sternberg, 2006).

permet de trouver de nouvelles idées sur des problèmes; Alors que la convergence mentale permet de synthétiser des idées disparates dans une solution nouvelle et appropriée. De toute évidence, les deux capacités sont utiles pour innover et doivent se coordonner quand de nouvelles idées viennent à l'esprit. Cependant beaucoup de nouvelles idées viennent à l'esprit tout le temps, mais si la personne est concentrée dans une direction spécifique, elle pourrait alors, par chance, attirer son attention sur un signal particulier et le convertir en une idée précieuse et créative. Ce qui semble indiquer une forme de divergence mentale, c'est-à-dire trouver de nouvelles idées, nécessitent aussi une forme de convergence mentale. Les choses qui viennent à l'esprit ne sont pas automatiquement interprétées comme des signaux: la plupart d'entre eux seront probablement rejetés comme du bruit et oubliés, et seules des idées ciblées appropriées seront conservées. Ainsi, la divergence et la convergence sont complémentaires dans la personnalité créative.

Intégrer l'évaluation du potentiel créatif dans les modèles et les théories économiques reste encore un grand défi pour deux raisons. Tout d'abord, il faut du temps pour mesurer la créativité, et les mesures sont mieux réalisées au sein du laboratoire ou dans tout autre environnement contrôlé, comme une salle de classe. Les mesures les plus fiables et complètes utilisent l'approche basée sur la production, dans laquelle les individus sont invités à produire un travail dans un domaine créatif donné. Une comparaison avec la production de d'autres personnes fournit une mesure de créativité (voir par exemple [Charness and Grieco, 2013](#)). Deuxièmement, l'évaluation dépend du domaine des productions créatives (graphiques ou verbales) et des modes de pensée (divergente ou convergente). Un exemple fiable et complet de l'évaluation du potentiel créatif chez les enfants à l'école est la batterie EPoC *Evaluation du Potentiel Créatif* de [Lubart, Besançon and Barbot \(, 2011\)](#), qui mesure le potentiel de la pensée divergente et convergente dans deux domaines différents, à savoir les arts graphiques et les compétences littéraires (verbal literacy). Cette procédure, qui est utilisée dans cette thèse, a

une grande validité: les auteurs ont trouvé une corrélation élevée et significative entre la pensée divergente et le test traditionnel de Torrance de la pensée créative (Torrance, 1962). En outre, ils ont trouvé une corrélation entre les mesures de créativité de l'EPoC et le trait de personnalité de l'ouverture, en accord avec les observations de McCrae and Costa (1987) selon lesquelles l'ouverture aux nouvelles expériences facilite la pensée divergente. La batterie complète du test EPoC dure environ deux heures.

2.3. Motivation et effort

La motivation correspond à l'ensemble des facteurs internes et externes qui stimulent les agents à faire un effort pour atteindre un but. Ainsi, la motivation a un impact important sur le comportement, y compris les décisions, les performances et les résultats. La motivation peut expliquer pourquoi les agents ayant des capacités différentes lors une tâche donnée atteignent le même résultat, et d'autre part, pourquoi les personnes ayant des capacités identiques ont des résultats différents.

La littérature distingue deux types de motivation, à savoir la motivation intrinsèque et extrinsèque. La motivation intrinsèque est motivée par un intérêt personnel ou une jouissance dans la tâche elle-même, généralement associée à l'apprentissage et à la créativité de haute qualité. La motivation extrinsèque provient d'influences externes afin d'atteindre un résultat souhaité, normalement caractérisé par des récompenses et des pénalités. Les motivations intrinsèques et extrinsèques sont liées à la performance, à la satisfaction, à la confiance et au bien-être (Gagné and Deci, 2005). Cependant, l'efficacité de la motivation extrinsèque pour promouvoir un effort durable est controversé parce que, dans certaines circonstances, elle pousse les agents à agir avec résistance et désintéressement pour atteindre les objectifs imposés (Deci and Ryan, 1985, 2010). L'impact de la récompense sur la motivation intrinsèque est également controversé. Pendant longtemps, le consensus dans la recherche en la psychologie sociale a mis en évidence un impact négatif des récompenses sur la motivation intrinsèque, et donc

sur la créativité. Alternativement, des études récentes montrent que, dans certaines circonstances, les récompenses améliorent la motivation extrinsèque sans détériorer la motivation intrinsèque (Gagné and Deci, 2005, Charness and Grieco, 2013), de manière équivalente Hennessey and Amabile (2010) déclarent que:

"[...] l'attente de récompense peut parfois augmenter les niveaux de motivation extrinsèque sans avoir d'impact négatif sur la motivation intrinsèque ou performance".

Compte tenu de l'effet ambigu de la récompense extrinsèque sur la motivation intrinsèque, la théorie de l'autodétermination (Deci and Ryan, 1985, 2010) distingue divers types de motivations extrinsèques, dont certaines représentent des formes passives de motivation et dont certaines représentent des formes de motivation actives (Ryan and Deci, 2000).

Les concepts de motivation autonome et contrôlée découlent de cette distinction. La motivation contrôlée est motivée par des expériences de pression et d'obligation, limitant le comportement souhaité à la période où la régulation externe est présente. Au contraire, la motivation autonome, également connue sous le nom de motivation durable, satisfait les besoins humains de compétence et d'autonomie. Elle est conforme à la motivation intrinsèque, fournissant aux individus le sens du choix, de la volonté et de l'autodétermination (Stone, Deci and Ryan, 2009). Ainsi, la motivation autonome peut avoir un impact important sur les résultats scolaires. Gagné and Deci (2005) suggère que:

"[...] parce que beaucoup des tâches que les professeurs veulent que leurs élèves effectuent ne sont pas intrinsèquement intéressantes ou agréables, savoir promouvoir des formes plus actives et volitives (contre passives et contrôlantes) de la motivation extrinsèque devient une stratégie essentielle pour réussir un enseignement".

Ici, nous mesurons la motivation par l'effort, c'est-à-dire la quantité de temps utilisée

pour effectuer une tâche d'effort réel dans le laboratoire, i.e. résoudre des anagrammes. Les variables de motivation sont liées à l'effort par définition (Brookhart, Walsh and Zientarski, 2006), donc une mesure d'effort est la meilleure approximation pour mesurer la motivation. Cependant, il est toujours complexe d'avoir une mesure précise de l'effort à l'école ou sur lieu de travail (Taylor and Taylor, 2011). Afin de combler cette lacune, les psychologues ont développé plusieurs échelles auto déclarées pour évaluer la motivation quand il n'est pas possible d'effectuer une mesure précise de l'effort. Ces échelles sont basées sur d'autres variables liées au concept de motivation, telles que l'estime de soi, l'auto-efficacité, l'autorégulation, le lieu de contrôle et l'orientation des objectifs. Par exemple, l'échelle Academic Motivation Scale (Vallerand et al., 1992) est développée pour mesurer la motivation intrinsèque et extrinsèque dans l'éducation.

Cette section a présenté l'importance des compétences non cognitives pour les résultats économiques et a fourni une revue de la littérature sur les compétences non cognitives abordées dans cette thèse. Il est important de noter que ces compétences peuvent être développées au cours du cycle de vie, mais leurs développements dès l'enfance est, de façon importante, déterminés par les caractéristiques sociales de la famille. Ainsi, selon une solide recherche, ces compétences peuvent expliquer (au moins partiellement) la persistance des inégalités en matière d'éducation. La section suivante traite de l'efficacité des systèmes scolaires, une autre source potentielle de biais sociaux - qui est abordée dans la partie II de cette thèse.

3. Comment les systèmes scolaires influencent les décisions éducatives, les résultats scolaires et la mobilité intergénérationnelle?

Les tests en mathématiques, en sciences et en lecture dans le Programme d'évaluation internationale des élèves (PISA) montrent que le niveau de réussite moyen des élèves

varie considérablement d'un pays à l'autre. Wößmann (2016) soutient que les différents systèmes scolaires sont responsables de façon considérable de la variation des performances entre les pays. Chaque pays a son propre système scolaire, qui comprend un ensemble de systèmes d'enseignement et enseignants formés par les politiques publiques. La comparaison entre pays montre que les systèmes d'enseignement tels que le suivi et le classement scolaire ont un impact important sur les décisions et sur les résultats des élèves. Par exemple, Wößmann (2016) montre que la compétition précoce dans différents types d'écoles par niveau scolaire augmente les inégalités en matière d'éducation, sans augmenter les niveaux de réussite¹⁸. La question est donc d'identifier quels sont les systèmes d'enseignement les plus efficaces.

3.1. Qu'est-ce qu'un système scolaire efficace?

Le concept d'efficacité est assez obtus pour les systèmes scolaires. Dans cette thèse, l'efficacité des systèmes scolaires implique un état économique dans lequel chaque ressource est répartie de façon optimale, en répondant de manière optimale à chaque agent. En d'autres termes, un système scolaire efficace maximise les résultats scolaires des élèves. De manière équivalente, un système scolaire efficace devrait minimiser les regrets espérés¹⁹, ce qui est associé aux échecs scolaires et aux élèves décrocheurs. Les étudiants qui ont échoué et qui ont subi une perte d'opportunité regretteront leur choix ex post et demanderont une redistribution en terme de politique publique. Alors que les politiques éducatives visent à réduire les échecs - considérés comme un gaspillage

18. Une des comparaisons expérimentales de l'efficacité des différents systèmes scolaires est l'un des objectifs de cette thèse. Il est présenté au chapitre 3

19. Dans la théorie de l'utilité espérée (UE), on sait que pour tout A, B : $EU(A) - EU(B) = EOL(B) - OL(A)$, avec $EOL(A)$ désignant la perte d'opportunité prévue de A par rapport à B (voir Raiffa, 1968, par exemple). Si $EOL(A)$ est la mesure du regret attendu de choisir A et précédant B . Donc, maximiser l'UE équivaut à minimiser le regret attendu. Bien que les deux programmes soient équivalents par la dualité, il est plus fréquent de parler de la maximisation de l'UE. Dans le contexte des politiques éducatives, le regret ex-post nourrit des frustrations et un mécontentement politiques, il est donc intéressant de le mentionner et de l'étudier. Cependant, les deux objectifs: maximiser l'UE et minimiser le regret attendu donnent les mêmes conclusions.

ou une inefficacité ex post -, les étudiants poursuivent leur propre objectif qui est de maximiser l'utilité espérée (UE) qui, malheureusement, ne garantit pas le succès futur et peut causer des regrets ex post.

Qu'est-ce qui guide les décisions éducatives? Le choix n'est pas une tâche psychologiquement simple. Les décisions importantes telles que les décisions éducatives²⁰, ou encore des décisions plus banales, peuvent entraîner des inquiétudes et surtout des inquiétudes concernant les occasions manquées causant un mécontentement même avec de bonnes décisions (Schwartz, 2004). Par conséquent, les élèves doivent apprendre à choisir afin de minimiser le regret.

Les décisions éducatives suivent les hypothèses d'utilité espérée. Les agents sont censés peser les perspectives d'une décision donnée avec leur probabilité de succès et sélectionner l'alternative qui leur fournira la plus haute utilité attendue. De manière équivalente, Heckman, Humphries and Veramendi (2016) affirment que:

"[...] dans le monde moderne, les individus devraient continuer leur scolarité tant que leur rendement marginal ex ante dépasse leur coût d'opportunité marginal ex ante".

Ici, il est important de postuler que, même si une décision éducative donnée est efficace ex ante, car elle maximise l'utilité attendue (ex ante), elle peut être inefficace ex post causant des regrets car il y a un risque croissant d'échec plus le niveau d'éducation est élevé.

Dans les sociétés modernes, les décisions éducatives sont probablement les choix les plus importants auxquels sont confrontés les individus au cours de leur vie. Elles sont importantes parce que la scolarité a une forte influence sur les résultats monétaires et non monétaires (Heckman, Humphries and Veramendi, 2016). L'effet positif

20. les élèves doivent décider de la prolongation de leur formation, c'est-à-dire décidé de soit de faire une nouvelle année d'études soit d'aller sur le marché du travail. Les élèves décident également de la filière, générale (ou académique), professionnelle ou technique. La variété des filières et le calendrier de ces décisions varient selon les systèmes scolaires.

de l'éducation va au-delà des résultats pécuniers, ce qui affecte également les comportements futurs, tels que le comportement de santé, le tabagisme, la consommation de drogues, la fertilité, la gestion du ménage, l'épargne, etc. (voir Vila, 2000, Lance, 2011, Król, Dziechciarz-Duda et al., 2013, Heckman, Humphries and Veramendi, 2016, entre autres).

Les choix éducatifs sont probablement les décisions les plus difficiles, l'éducation est donc le domaine de vie comportant le plus grand potentiel de regret possible dans la société contemporaine (Roese and Summerville, 2005). Ces choix éducatifs sont difficiles pour deux raisons principales. Tout d'abord, comme nous l'avons vu dans le paragraphe précédent, l'éducation a des conséquences monétaires et non monétaires importantes pour l'avenir des individus. Par conséquent, estimer un retour à la scolarité est une décision et un défi complexe pour les agents. Deuxièmement, parce que ces décisions sont entourées d'une grande incertitude, puisque les élèves ont une connaissance imparfaite de leur capacité et de leurs préférences lorsqu'ils décident. Par conséquent, l'environnement familial et social joue un rôle important et influence fortement les choix éducatifs.

Qu'est-ce qui prédit la probabilité normative de succès à l'école? Dans le contexte éducatif, la perspective de salaires élevés après avoir suivi des études dans l'enseignement supérieur peut pousser les élèves à essayer rationnellement d'accéder à niveaux d'éducation plus élevés, même pour ceux qui ont peu de chances d'y réussir. Cependant, les agents rationnels tiennent compte de leur probabilité de succès avant de prendre leur décision afin d'éviter les échecs et les regrets, de sorte qu'une bonne estimation des chances futures de réussite est cruciale pour prendre des décisions optimales. La question se pose alors de *qu'est ce qui prédit le succès à l'école?* Les progrès récents dans la littérature économique affirment que la réussite scolaire dépend d'un ensemble étendu de compétences cognitives et non cognitives, telles que: la persistance de la motivation, l'estime de soi, la tolérance au risque, l'optimisme et préférences

de temps (voir Heckman and Rubinstein, 2001). Par exemple, Castillo et al. (2011) ont constaté que plus les enfants sont patients, plus ils s'intéressent aux conséquences futures de leur comportement, et plus ils ont des perspectives favorables.

Il n'y a pas de consensus sur le fait que l'intelligence, mesurée par le QI, est un prédicteur important du succès scolaire. Un QI plus élevé augmente la performance à l'école, mais ce n'est pas une garantie de réussite future si l'élève n'est pas motivé dans ses études. La motivation de la réussite est également importante pour la réussite scolaire (Busato et al., 2000) et a un impact direct sur la perception des élèves selon laquelle le succès dépend de l'effort de chacun (Ames and Archer, 1988). L'effort (et la motivation) peut expliquer pourquoi les élèves ayant des niveaux de capacité cognitive différents peuvent atteindre le même résultat scolaire.

Dans la même veine, nous ne pouvons pas négliger que les notes réelles sont expliquées autant par des capacités cognitives et non cognitives. Les résultats scolaires peuvent donner aux élèves des signaux importants quant à leurs performances futures. Le problème de la prise de décision sur les résultats scolaires est que cette mesure ne tient pas compte du niveau croissant de difficultés dans l'éducation. En outre, il ne garantit pas la motivation future des élèves si ces derniers ne choisissent pas leur cursus futurs selon leurs préférences. Selon Schwartz (2009), un élève ayant des doutes quant à savoir s'il a fait le bon choix éducatif peut probablement être moins engagé dans ses études que quelqu'un qui ne doute pas. Ainsi moins d'efforts se traduisent vraisemblablement par une mauvaise performance.

En résumé, même s'il existe des indices importants pour prédire la réussite scolaire future, il est peu probable que les élèves puissent estimer correctement les chances normatives de succès lorsqu'ils font leurs choix éducatifs. Ici, il est important de postuler que, comme les élèves ne connaissent pas ex ante leur véritable probabilité de succès pour un autre cursus scolaire, les décisions éducatives sont basées sur leur probabilité subjective de réussir dans cursus donné, c'est-à-dire leur confiance en soi.

Compte tenu de la complexité nécessaire pour comprendre les chances de réussite scolaire future, les politiques publiques peuvent jouer un rôle important en aidant les élèves à faire correspondre leurs capacités à des décisions de choix optimales, et donc à accroître le bien-être et à diminuer les regrets. [Filippin and Paccagnella \(2012\)](#) présentent un exemple intéressant de ce genre de politique éducative lors de la comparaison des systèmes scolaires néerlandais et italien. La principale différence entre ces deux pays concerne l'auto-sélection pour les filières du lycée²¹. Aux Pays-Bas, les résultats d'un test d'aptitude national à l'âge de 12 ans indiquent la filière la plus appropriée aux élèves selon leurs aptitudes. Au contraire, les élèves - et leurs parents - n'ont aucun signal pour choisir la filière appropriée en Italie. La figure II.1 montre l'efficacité de la capacité de signalisation des étudiants aux Pays-Bas. Le faible degré de chevauchement entre les trajectoires éducatives au Pays-Bas suggère une meilleure correspondance entre la capacité et la filière éducative choisie lorsque les élèves ont de bons signaux sur leurs capacités.

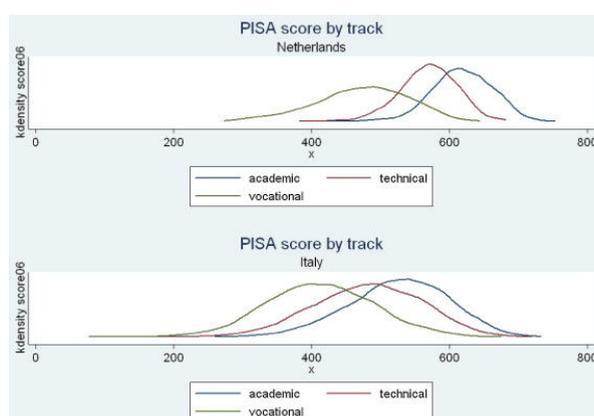


Figure II.1 – Suivi du cursus secondaire par aptitude. Réimprimé de [Filippin and Paccagnella \(2012\)](#) avec la permission.

Dans le même ordre d'idées, [Goux, Gurgand and Maurin \(2016\)](#) montrent une expérience aléatoire contrôlée en France, dans lequel les étudiants en difficulté scolaire

21. Les trois filières possibles ici sont: académique (générale), professionnelle et technique.

et leurs familles ont eu plusieurs rencontres avec les directeurs de collège. Le but de ces réunions était d'expliquer: (i) l'importance des choix qu'ils devraient faire avant la fin de l'année scolaire²², et (ii) que la performance réelle des élèves devait être plus importante pour qu'elle s'ajuste aux décisions éducatives auxquelles les parents aspiraient. Ce programme a aidé les élèves (et les familles) à formuler des objectifs éducatifs mieux adaptés à leurs aptitudes académiques, remettant en question les décisions de poursuite scolaire au lycée pour les étudiants les moins réalistes. Par conséquent, ce programme a réduit les échecs à l'école secondaire de 25% dans la population cible.

Cette section montre que, en plus d'avoir de plus grandes chances de réussite, les étudiants les plus capables sont susceptibles d'avoir des choix éducatifs les plus appropriés. Ces deux observations peuvent donc être l'une des causes de l'inégalité observée dans la réussite scolaire, qui est discuté dans la section suivante.

3.2. Comment réduire l'écart social dans les résultats scolaires?

L'importance de l'éducation pour la mobilité intergénérationnelle fait consensus pour les spécialistes des sciences sociales (par exemple [Piketty, 2000](#), [Black and Devereux, 2011](#)). Ainsi, la compréhension des causes de l'inégalité dans la réussite scolaire est une question cruciale pour promouvoir la mobilité intergénérationnelle.

Biais socioéconomique Nous définissons le biais socioéconomique dans l'éducation dans la mesure où les décisions éducatives et le niveau de scolarité sont influencés par le statut socioéconomique des élèves. Selon cette hypothèse, plus le système éducatif est favorable à un groupe social donné, plus le système socio-économique est biaisé. Ainsi, si les enfants des classes socio-économiques supérieures sont surreprésentés dans les niveaux d'éducation supérieurs alors cela constitue une preuve de biais socioéconomique dans le système éducatif.

22. en France, les étudiants doivent décider de la filière du secondaire (académique ou professionnelle) à la fin du lycée.

L'analyse classique de [Becker \(1967\)](#) sur la mobilité intergénérationnelle attribue l'inégalité des chances essentiellement aux différences de capacités et d'opportunités. Étant donné que les capacités innées sont également réparties dans toutes les classes sociales, le biais socioéconomique de l'éducation disparaît une fois que les différences sociales dans les opportunités éducatives peuvent être neutralisées. La théorie du capital humain ([Becker, 1964](#)) démontre que la présence d'un marché de crédit efficace concernant les investissements dans l'éducation est tout ce qui est nécessaire pour atteindre cet objectif. Cette prédiction optimiste n'a pas été tout à fait concrétisée dans les pays développés, malgré des efforts soutenus pour éradiquer les différences dans les opportunités éducatives entre les différentes classes sociales. Plusieurs études ont montré que les différences dans les opportunités éducatives ne représentaient qu'un rôle négligeable dans les pays développés ([Carneiro and Heckman, 2002](#), [Cameron and Taber, 2004](#)).

Par conséquent, la persistance des inégalités dans le niveau de scolarité démontre la présence de capacités non cognitives socioéconomiques biaisées, accumulées pendant l'enfance et l'adolescence. Ces capacités non cognitives sont héritées par les enfants et les jeunes de leur exposition permanente à leurs parents, amis, pairs et environnement social et de l'investissement différentiel des familles dans leur capital humain ([Lévy-Garboua, 1973](#), [Becker and Tomes, 1979](#), [Cunha and Heckman, 2008](#), [Heckman and Farah, 2009](#), [Cunha, Heckman and Schennach, 2010](#)). Il ne fait aucun doute que les enfants de classes supérieures sont susceptibles de grandir dans de meilleurs environnements d'apprentissage, avec plus de stimuli²³ et moins de stress ([Heckman, 2011](#)). En outre, les enfants de statut socioéconomique différent n'ont pas les mêmes choix éducatifs car ils n'ont pas les mêmes points de référence et les mêmes niveaux d'aspiration: les enfants de SES inférieures peuvent considérer comme un succès ce que les personnes de SES plus élevées considèrent comme un échec ([Boudon, 1973](#)). James Heckman utilise

23. Par exemple, le nombre de livres à la maison est le prédicteur le plus important du rendement scolaire ([Fuchs and Woessmann, 2004](#)).

le terme "*loterie de naissance*" lorsqu'il décrit l'effet puissant de l'héritage de la famille dans la façon de façonner la trajectoire de sa vie. En effet, plusieurs études affirment que les caractéristiques de base de la famille sont plus importantes que la composition sociale de l'école et les ressources scolaires pour prédire les résultats scolaires (Chudgar and Luschei, 2009, Borman and Dowling, 2010).

Considérer le développement de compétences non cognitives dans l'éducation infantile peut être un facteur d'égalité dans la compétition sélective pour une position sociale. Les écoles doivent cibler le développement infantile des capacités non cognitives afin de réduire l'effet des antécédents familiaux dans la mobilité intergénérationnelle. Un système équitable devrait améliorer les résultats des individus en difficulté, sans préjugés des individus qui ont plus de facilité.

Nous avons vu que les capacités non cognitives sont importantes pour la réussite de la vie et peuvent avoir un impact notable sur les décisions et les performances en matière d'éducation. C'est une explication possible de la persistance des inégalités dans les résultats scolaires. Cette thèse vise à étudier trois aptitudes non cognitives, à savoir la créativité, la confiance en soi et la motivation et l'impact de ces deux dernières sur les décisions éducatives et les inégalités en matière d'éducation. La section suivante présente les méthodes expérimentales (utilisées dans la thèse) et montre en quoi elle convient parfaitement à cette approche.

4. L'utilisation de méthodes expérimentales pour étudier les capacités non cognitives et les établissements scolaires

La mesure des capacités non cognitives est un défi pour les économistes (Humphries and Kosse, 2017). L'évaluation de la plupart des variables psychologiques repose sur des tests et des enquêtes nécessitant du temps et, de préférence, un environnement con-

trôlé. Ainsi, l'utilisation d'expériences de laboratoire semble l'alternative la plus appropriée pour introduire des variables psychologiques dans la recherche économique²⁴. L'expérience de laboratoire nous permet de mesurer des variables qui sont au cœur de cette thèse, comme la créativité, la performance, la capacité, la confiance en soi et les efforts qui seraient difficiles à observer précisément dans les enquêtes.

Le principal avantage des expériences en laboratoire est la possibilité d'isoler des variables d'intérêt spécifiques tout en contrôlant l'environnement. Ce mécanisme permet l'isolement et l'identification des effets causaux. Dans ce contexte, les méthodes expérimentales sont un outil puissant pour tester les théories, rechercher de nouveaux faits, comparer les institutions et les environnements et tester les politiques publiques.

En général, les environnements créés en laboratoire sont plus simples que ceux qui se trouvent dans la nature. La question de la validité externe des expériences de laboratoire se pose, c'est-à-dire dans quelle mesure le comportement en laboratoire est-il corrélé au comportement de la vie réelle et les résultats d'une étude peuvent-ils être généralisés? Ce problème est controversé parmi les économistes. [Levitt and List \(2007\)](#) résumant les principales critiques concernant la validité externe des mesures en laboratoire. Les auteurs affirment que dans le laboratoire: (i) le contexte, les ensembles de choix et les horizons de temps ne peuvent pas être complètement répliqués dans le laboratoire, (ii) les caractéristiques des sujets expérimentaux diffèrent des groupes impliqués dans des décisions hors laboratoire, (iii) les incitations monétaires sont différentes de la vie réelle. Dans une réponse critique à [Levitt and List \(2007\)](#), [Camerer \(2015\)](#) fournit une perspective plus favorable pour l'expérience, il soutient que: (i) la validité externe n'est pas une préoccupation principale dans une expérience type, puisque l'économie expérimentale vise à établir une théorie générale reliant les facteurs économiques tels que les incitations, les règles et les normes aux décisions et au comportement, (ii) certaines expériences ont des fonctionnalités qui ne peuvent

24. En fait, l'économie expérimentale en laboratoire est inspirée des expériences en psychologie sociale.

pas être généralisées sur le terrain, comme certains paramètres de champ qui ne peuvent être généralisés à d'autres paramètres de champ, (iii) La plupart des expériences économiques examinées par cet auteur sont résumées dans cet article, et montrent une corrélation entre le comportement en laboratoire et hors laboratoire. Dans le même ordre d'idées, [Plott \(1991\)](#) soutient que dans le laboratoire:

"[...] les vraies personnes motivées par de l'argent réel prennent des décisions réelles, commettent des erreurs réelles et subissent de réelles frustrations causée par leurs vrais talents et de leurs vraies limites".

Ainsi, même s'il n'est pas possible de reproduire des environnements naturels, nous pouvons avoir des indices valables à partir des comportements et de décisions prises en situation expériences incitatives de laboratoire.

Les quatre chapitres de cette thèse présentent des expériences de laboratoire incitatives. Les chapitres [1](#) et [2](#) sont basés sur des tâches d'effort réel où la prise de décision et le comportement sont observés et analysés afin d'étudier la confiance en soi et la créativité. Les chapitres [3](#) et [4](#) proposent un système d'enseignement type expérimental qui permet de comparer les différents systèmes scolaires et les établissements d'enseignement.

Il existe plusieurs contraintes empiriques pour étudier un contexte institutionnel donné et / ou faire des comparaisons internationales à l'aide de données sur le terrain car il est presque impossible d'isoler l'effet étudié en maintenant tout le reste. L'utilisation d'un cadre expérimental est une bonne alternative pour étudier les établissements d'enseignement, même s'il n'est pas possible de saisir tous les éléments d'un système éducatif dans un environnement contrôlé. Pour aborder cet objectif de recherche, nous reproduisons au moyen d'une expérience de laboratoire incitatif le système éducatif type ²⁵, où nous pouvons discriminer selon: les différences dans les programmes scolaires, les différences de rémunération, le choix de la filière et le niveau

25. Le système éducatif type utilisé dans cette recherche est décrit en page [127](#).

de performance nécessaire pour être admissible à un certain programme d'études.

Les résultats expérimentaux permettent une comparaison simple et valable de la performance globale d'un système scolaire minimal dans différents mécanismes de tri, ce qui facilite l'identification d'un design efficace, le design maximal de la production éducative conditionnel à la distribution de la capacité.

5. Aperçu de la thèse

Cette thèse présente deux parties principales, respectivement organisées en deux chapitres. La première partie est consacrée à étudier deux compétences non cognitives impliquées dans le processus de décision, qui sont la confiance dans le succès futur et le potentiel de la créativité. La confiance en soi sur le succès futur est un sujet d'intérêt pour les économistes depuis longtemps, car il a été un déterminant important du processus de décision: on suppose que les individus maximisent leur utilité attendue en fonction de leur probabilité (subjective) des différents résultats auxquels ils font face. Si, d'une part, la confiance en soi est largement étudiée par les économistes, d'autre part, le potentiel de créativité est une nouvelle variable d'intérêt dans ce domaine de recherche. La mesure psychologique du potentiel de la créativité évalue dans quelle mesure un individu peut s'engager dans un travail créatif (Lubart, Zenasni and Barbot, 2013) et peut donc avoir un impact important sur les résultats économiques. La deuxième partie de cette thèse se concentre sur l'étude expérimentale des systèmes scolaires, avec deux objectifs principaux: la comparaison de l'efficacité des différents mécanismes de tri scolaire et une évaluation des biais sociaux et de genre qui en découlent.

Le chapitre 1 compare la vitesse d'apprentissage de sa capacité spécifique dans un jeu double ou à quitter avec la rapidité de la confiance croissante à mesure que la tâche devient de plus en plus difficile. Nous constatons que les gens en moyenne apprennent à être trop confiants plus rapidement qu'ils apprennent leur véritable capacité et nous présentons un modèle de confiance intuitif-bayésien qui intègre ces faits. L'incertitude

quant à sa capacité réelle à effectuer une tâche en isolement peut être responsable de biais de confiance importants et stables tels que: une discrimination limitée, l'effet de la difficulté, l'effet Dunning-Kruger, l'apprentissage conservateur de l'expérience et le phénomène de dépassement (sans précisions) si les sujets agissent comme apprenants bayésiens (i.e qui ne dépendent que de signaux de performance perçus séquentiellement et de signaux illusoire contraires induits par le doute). En outre, ces biais sont susceptibles de persister si l'agrégation bayésienne de l'information antérieure consolide l'accumulation d'erreurs et la perception des signaux illusoire contradictoire génère un conservatisme et une sous-réaction aux événements. Ensemble, ces deux caractéristiques peuvent expliquer pourquoi les «Bayésiens» intuitifs font systématiquement de mauvaises prédictions de leur propre performance.

Le chapitre 2 vise à comprendre l'impact de la créativité sur les résultats économiques. Le premier objectif de ce chapitre est d'examiner comment les économistes décrivent le comportement créatif et proposent comment il devrait être décrit. Nous soutenons que, du point de vue économique, le comportement créatif doit être jugé par la propension à innover dans les activités de production (et de consommation), en distinguant deux types d'innovateurs économiques: les *chercheurs* (la capacité de trouver de nouvelles solutions) et les *entrepreneurs* (la capacité de produire efficacement le travail avec la technologie existante). Le deuxième objectif est d'observer comment le potentiel de la créativité influence la production des individus. Nous proposons une expérience économique avec deux tâches d'effort réel pour observer la performance des individus créatifs en production, en utilisant trois mesures psychologiques de la créativité: la pensée divergente graphique, la pensée intégrative graphique et l'indice de créativité agrégé. Nous trouvons que la pensée divergente est en corrélation avec le type d'innovateur économique du chercheur puisque les scores plus élevés pour cette mesure psychologique de la créativité augmentent la productivité des activités d'exploration. Cependant, le type entrepreneurial n'a pas été identifié parmi nos scores de créativité.

En outre, nous observons que les individus créatifs ne sont pas plus productifs que d'autres dans des tâches répétitives, mais ils se comportent différemment que des individus moins créatifs dans ce type de tâche: les penseurs intégratifs sont plus coopératifs lorsqu'ils travaillent par paires, peut-être parce qu'ils sont intelligents et comprennent l'avantage de la coopération dans le travail d'équipe. En effet, l'idée que les individus créatifs sont intelligents est renforcée par la performance à l'école - une performance de la vie réelle. Les scores de créativité jouent un rôle important dans les réalisations scolaires, ils sont corrélés positivement aux notes sur les mathématiques, le français et la moyenne générale.

Les chapitres 3 et 4 sont basés sur la même expérience et ensemble de données, avec un total de 941 participants. Nous reproduisons expérimentalement l'archétypique de la structure des systèmes scolaires et proposons une tâche d'effort réel (résolution d'anagrammes). Après une longue phase de scolarité obligatoire (niveau 1), les étudiants peuvent quitter le marché du travail ou entreprendre d'autres études. Ceux qui décident de continuer ont habituellement une option entre deux filières (ou plus), une filière générale et une filière professionnelle, qui diffèrent selon le niveau requis de capacité cognitive. Les étudiants moins capables devraient opter pour des études professionnelles au niveau 2, alors que les plus capables opteraient pour des études générales. Si elle réussissait, les deux groupes d'étudiants auraient un autre choix pour quitter ou entreprendre d'autres études (niveau 3). Cependant, les étudiants engagés dans l'enseignement général devraient généralement trouver plus facile de passer ce niveau supérieur que les étudiants engagés dans une carrière professionnelle. Nous comparons quatre mécanismes pour trier les élèves en fonction de leurs capacités: auto-sélection d'études ultérieures avec une seule filière (sans choix de filière), auto-sélection d'autres études avec le choix de la filière (choix), le dépistage par aptitude et la compétition.

Le chapitre 3 montre que le choix et le dépistage sont des mécanismes plus efficaces, offrant des gains, des résultats et un taux de réussite plus élevés au niveau universitaire.

Le dépistage donne le résultat le plus élevé (nombre d'anagrammes résolus) pour le niveau primaire, car il stimule l'effort soutenu des individus à ce niveau. La concurrence anticipée («compétition») est le pire traitement, car les participants s'occupent non seulement de leur propre performance, mais aussi de la performance des autres. Le problème de l'auto-sélection («choix») est qu'il favorise le plus haut niveau d'échec au niveau secondaire lorsque le retour, le coût économique à l'école est élevé. En fait, nous observons que les retours à l'enseignement supérieur augmentent le montant des regrets. Ainsi, nous observons que l'inefficacité du système découle de deux raisons principales: (i) si les élèves ont une connaissance imparfaite de leur propre capacité et / ou s'il existe un manque de discrimination entre les deux filières éducatives (formation professionnelle et générale), ils sont enclins à opter pour la filière et l'échec les plus difficiles, ce qui entraîne une perte de bien-être de 12% par rapport à l'expérience; (ii) le plus élevé est la prime de salaire pour l'enseignement supérieur, plus l'utilité anticipée est ex ante, ce qui augmente les chances d'essayer des niveaux d'éducation plus élevés, mais n'accroît pas la probabilité de succès, ce qui augmente l'inefficacité ex post (Niveau supérieur d'échecs, d'abandon et de regret).

La question étudiée dans le chapitre 4 est: *comment les différents systèmes scolaires et les résultats scolaires affectent-ils différemment les groupes de capacités (de niveaux), le genre et les groupes sociaux, ce qui entraînent des différences substantielles dans les biais sociaux et genrés entre les pays développés en différentes périodes?* Nous constatons que la *compétition* est la pire institution pour les personnes ayant une capacité élevée et moyenne, tandis que l'auto-sélection de la piste est le pire traitement pour les personnes à faible capacité lorsque les retours à l'enseignement supérieur sont élevés. Le principal résultat observé lors de l'augmentation des bénéfices pour le niveau tertiaire est que l'augmentation de l'enseignement supérieur est bénéfique pour les étudiants à haute et moyenne capacité, mais elle est nocive pour les élèves à faible capacité. La comparaison des bénéfices pour les traitements à deux choix semble injuste

au début, mais étonnamment, les participants à faible capacité ont gagné 22% de plus dans la condition avec des incitations inférieures. Cet effet est dû à de meilleures décisions.

Ainsi, étant donné que la concurrence est particulièrement dangereuse pour les individus à haute et moyenne capacité, cela semble entraîner un biais social plus faible que les autres mécanismes de tri, conséquence directe de l'inefficacité relative de ce mécanisme. L'impact de l'auto-sélection (Choix) avec des incitations élevées sur la performance des participants moins capables a l'effet contraire, il génère le biais social le plus élevé parmi nos traitements. Nous observons également que l'allocation aléatoire est le seul mécanisme équitable pour les différences de genre. Le dépistage semble être le mécanisme le plus équilibré pour suivre les étudiants par capacité, le défi est de définir des notes justes (seuils) qui encouragent les étudiants motivés à faible et moyenne capacité à atteindre des niveaux de scolarité plus élevés, sans décourager les moins motivés pour compléter le niveau primaire. Un système équitable devrait améliorer les résultats des individus moins capables, sans préjugés pour les plus capables.

Résumé

Cette thèse contribue à la littérature économique sur les compétences *non cognitives* essentielles pour la réussite de la vie, particulièrement pour la réussite scolaire. Elle comprend quatre essais basés sur des approches économiques comportementales et expérimentales, avec deux objectifs principaux. Le premier objectif est d'étudier deux compétences non cognitives, à savoir la confiance en soi et la créativité. Notre but est alors de comprendre les déterminants de la confiance en soi et l'impact de la créativité sur les résultats économiques. Le deuxième objectif est d'étudier comment le système scolaire influence les décisions éducatives, les résultats scolaires et la mobilité intergénérationnelle, secteurs où les compétences non cognitives peuvent jouer un rôle important, en particulier via leurs effets sur la confiance en soi et la motivation. Nous observons un impact important des capacités non cognitives sur le comportement et sur les résultats économiques, notamment en ce qui concerne les décisions éducatives. Tout nous montre à penser que fondamentalement les écoles s'inquiètent du développement de ces capacités non cognitives - et non pas seulement des capacités cognitives. Aider les élèves à avoir de meilleures estimations de leur confiance en soi, favoriser le développement du potentiel créatif, stimuler la motivation et l'effort devraient alors faire partie de l'éducation que les élèves reçoivent dans les écoles; promouvant alors de meilleures décisions, de meilleurs résultats et une société plus équitable.

Abstract

This thesis contributes to the growing economic literature on *noncognitive* skills that are critical for life success, specially for academic success. It comprises four essays based on behavioral and experimental economics approaches, with two main objectives. The first objective is to study two noncognitive skills, namely self-confidence and creativity. We aim at understanding the determinants of self-confidence, and the impact of creative potential on economic outcomes. The second objective is to study how school systems impact educational decisions, educational outcomes and intergenerational mobility, where noncognitive skills may play an important role, specially self-confidence and motivation. We observe an important impact of the noncognitive abilities on behavior and economic results, especially for the educational achievements. Taken all our evidences together, it seems fundamental that schools worry about the development of these noncognitive abilities - and not only of the cognitive abilities. Helping students to have better estimates of self-confidence, favoring the development of creative potential, and stimulating motivation and effort should be part of the education that pupils receive in schools; promoting then better decisions, better outcomes and a more equitable society.

Keywords: noncognitive abilities, education, school systems, creativity, motivation, effort, self-confidence, overestimation bias, educational inequalities, intergenerational mobility.