Demographic Research a free, expedited, online journal
of peer-reviewed research and commentary
in the population sciences published by the
Max Planck Institute for Demographic Research
Konrad-Zuse Str. 1, D-18057 Rostock · GERMANY
www.demographic-research.org

DEMOGRAPHIC RESEARCH

VOLUME 23, ARTICLE 21, PAGES 587-614
PUBLISHED 14 SEPTEMBER 2010
http://www.demographic-research.org/Volumes/Vol22/21/
DOI: 10.4054/DemRes.2010.23.21

Research Article

Siblings and human capital: A comparison between Italy and France

Giulia Ferrari

Gianpiero Dalla Zuanna

This publication is part of the proposed Special Collection "Social Mobility and Demographic Behaviour: A Long-Term Perspective", organized by Guest Editors Cameron Campbell, Jan Van Bavel, and Martin Dribe.

Table of Contents

1	Introduction	588
2	Background, hypotheses, and data	589
2.1	Background	589
2.2	Hypotheses	591
2.3	Data	592
3	Descriptive analysis: Method and variables	593
3.1	Method	593
3.2	Response variable	593
3.3	Covariates	594
4	Descriptive analysis: Results	597
4.1	Testing the first hypothesis: The dilution effect holds in both countries, but is stronger in Italy	597
4.2	Testing the second hypothesis: The dilution effect is weaker among French bourgeois families	599
4.3	Testing the third hypothesis: No differences by cohort	600
4.4	Testing the fourth hypothesis: It is better to be the first-born child	602
5	Toward a causal approach	603
6	Discussion and conclusion	608
7	Acknowledgements	609
	References	610
	Appendix: Constructing the instrumental variable	612

Demographic Research: Volume 23, Article 21
Research Article

Siblings and human capital: A comparison between Italy and France

Giulia Ferrari[1]

Gianpiero Dalla Zuanna[2]

Abstract

This paper investigates how family size affects children's human capital, comparing Italy and France. We tested the dilution effect in these countries, starting with the assumption that the higher the number of siblings, the fewer parental resources are available for each child, and the lower the probability that each child will successfully pursue his/her educational career. We find a negative correlation between the number of siblings and human capital. However, when the analysis is developed with a causal approach, the strength of the dilution effect weakens in Italy and disappears in France.

[1] PhD Student of Demography, University of Rome "La Sapienza." E-mail: giulia.ferrari@uniroma1.it
[2] Professor of Demography, University of Padova.

1. Introduction

Studies of the determinants of educational attainment, occupation, and income usually consider the family background as a general effect. Meanwhile, fewer researchers have investigated family size as an independent effect. In this article, we focus our attention on a set of fundamental questions regarding social inequalities that may be linked to family background characteristics, especially family size. Based on the *dilution effect theory*, our assumption is that the higher the number of siblings, the lower the likelihood is that an individual will attain the highest levels of education (Blake 1980). Does growing up in large families affect people's educational chances? If so, is the negative effect on educational attainment of belonging to a large family weakened among the more advantaged social classes? Has the negative effect of large family size on educational attainment changed over the course of the 20th century? Are there any significant differences in the educational opportunities and life chances of first- and lastborns?

When families have large numbers of children, there is often a dilution of the resources available to each individual child. This research seeks to investigate whether, as a result of varying national family policies, the "dilution effect" differs between countries. In our analysis of this issue, we focus on Italy and France, two European countries where family policies and fertility choices are very different, but for which comparable data are available.

The main goal of this article is to describe the statistical association between siblings and human capital, after controlling for other variables (year at birth, social class, etc.). We note, however, that some unobserved factors may also influence both family size and children's outcomes. Previous studies (see, for example, Angrist and Evans 1998) have argued that one possible source of unobserved heterogeneity is the preference among parents to have smaller families, with each child being of higher *quality*. The level of "status anxiety" for children may drive *both* the family size *and* the education of each child (Dalla Zuanna 2007). Hence, the endogeneity issue should be taken into account when evaluating the *true causal effect* of family size on education. In the final section of this article, an instrumental variable approach will be applied, using the sibling sex composition as an exogenous determinant of family size variation (Conley and Glauber 2005). This approach – which was mainly chosen because of the specificity of the data source for Italy and France – is not completely satisfactory. Consequently, this part of the research should be viewed as merely a first step towards the formulation of a more exhaustive causal approach.

We intend to contribute to the literature published to date by providing the first international comparison in this area of research. It certainly is a challenge to compare two countries, especially when the surveys are not identical. Fortunately, the variables

of interest examined in this paper are quite similar for the two countries, enabling us to compare different behaviors, as well as different policies.

The paper is organized as follows. The next section briefly summarizes the current state of the debate about these issues in the literature, and outlines the main hypotheses. Sections 3 and 4 are devoted to a descriptive analysis of the association between siblings and human capital, using simple cross-tabulations and logistic models. In Section 5, the causal analysis is introduced and developed. Finally, Section 6 summarizes the results and suggests directions for further research.

2. Background, hypotheses, and data

2.1 Background

The level of interest among researchers in the relationship between family size and the educational attainment of the children in the family has increased in recent decades. One pattern that has frequently been observed in the literature is that as the number of siblings increases, the educational achievement of the children in the family declines.

This issue was studied in detail by Judith Blake in the 1970s. Blake's results were reported in her book, *Family Size and Achievement,* which was published in 1980. Blake coined the term "*dilution effect*," which refers to the increasing disadvantages encountered by individuals who grow up in large families. Blake asserts that, since the family's material and non-material resources are limited, children reared in larger families are more likely to suffer from less access to these resources. In addition, the differential allocation of resources to each child is assumed to affect educational outcomes, as well as intellectual development. It is further posited that the total amount of resources available depends on the number of children in the family, and on how they are spread out in age. In her book, Blake presents results for the U.S. that are based on a number of longitudinal and cross-sectional surveys on the relationship between educational attainment and the number of siblings, as well as on ethnicity, religion, and social status. Moreover, she explores the association between sibship size and intelligence, as well as the effect of birth order on educational outcomes and intellectual ability. She concludes that the reduction of family sizes experienced in the U.S. in recent decades may be expected to improve educational opportunities for individuals. In fact, evidence shows that, even after controlling for major parental background characteristics, the larger the sibship size, the lower the probability that children will achieve high levels of education, the higher the likelihood that children will drop out of school, and the lower the children's IQ.

In addition to family size, another crucial characteristic of the siblings group—namely, birth order—should be taken into account. The literature presents conflicting empirical results of analyses of the impact of birth order on educational attainment, with different studies arguing that either firstborns and lastborns can benefit from their rank position. On the one hand, some authors conclude that firstborn children are advantaged with respect to their later born siblings because of their priority in obtaining parental time, energy, and attention. On the other hand, these benefits do not generally extend to economic resources. In fact, later born children are more likely than the firstborn to have older parents, who in turn are more likely to be in a better financial position to support their children. In addition, later born children could exploit their older siblings' knowledge to obtain help in doing their homework, which could increase their probability of succeeding at school. Recognizing their disadvantageous position in the family hierarchy, later born children also develop alternative *strategies of survival that often entail risk-taking and daring behaviour* (Black, Devereux and Salvanes 2005). By contrast, other scholars (Gary-Bobo, Prieto, and Picard 2006) argue that a higher birth order confers no advantages, and that being a firstborn may even have a significant and negative impact on educational achievement. According to their findings, these negative effects remain even after controlling for the father's occupational status.

As our paper primarily focuses on the association between sib-size and education, we do not take into account other characteristics of the sibship, e.g., age spacing between children or the sex composition of siblings. Today, families may vary more in the age spacing of children than in the number of siblings. Scholars generally agree that longer intervals allow parents to recover economic losses before the next child requires further investments. Regarding the gender composition of siblings, findings vary widely and often appear to be inconsistent. For example, the father's occupation has been shown to have different effects for boys and girls, with some studies suggesting that fathers who are self-employed or in certain professions may encourage their sons—but not their daughters—to follow in their footsteps (Zarca 1995a,1995b).

This article relies on Judith Blake's *dilution theory*. One of the principal advantages of this theory is that it allows us to find evidence using a number of empirical applications (Steelman et al. 2002). However, Blake is not the only researcher who has sought to explain the relationship between family size, birth order, and age spacing. Thus, to provide a more complete research framework for this analysis, we cite the Zajonc and Marcus' *confluence theory*: "[A]ccording to this theory, the developing child is moulded by the intellectual atmosphere to which he/she is exposed in the family setting. (…) the intellectual climate is calculated by averaging the intellectual level of all members of the family" (Steelman et al. 2002). In other words, they challenge the perceived drawbacks of having more children by looking instead at the mean maturity

level of the siblings, but they are not able to explain the relationship between educational achievement and family size that is net of intellectual development.

In recent decades, scholars have become increasingly interested in the question of whether the association between family size and children's outcomes could represent a real causal influence (for a review, see Angrist, Lavy and Schlosser 2005). The observed relationship could be spurious because of the simultaneous determination of outcome and "treatment." Specifically, a considerable source of unobserved heterogeneity might arise from the parents' preference for having fewer children, but of higher *quality*. Couples who are very concerned about their children's education would try to keep the family size small in order to allow each child to achieve a high level of education. Many other factors may also influence both educational achievement and family size, such as such where the family lives and whether they are recent immigrants. Unfortunately, our data do not provide enough information to allow us to take these factors into account.

2.2 Hypotheses

Having taken these research findings into consideration, our general hypothesis overlaps with Judith Blake's: i.e., we assume that, as the number of siblings increases, the individual's opportunities of achieving high levels of education decrease. The so-called dilution effect is a concept very close to that of the quantity-quality trade off (for a detailed discussion of the conflict between having another child and the desire to sustain the social mobility of child(ren) ever born, see, for example, Dalla Zuanna 2007). By means of our data, this general hypothesis will be tested using the following steps.

(1) *The dilution effect is assumed to be more intense in Italy than in France* because the French instrument of *"quotient familial"* (a coefficient which operates by reducing the household's income tax according to the number of children), and the more family-friendly policy orientation of this country, should provide financial incentives for increasing births, while, at the same time, indirectly decreasing the burden of educational expenses. We expect that, when following a causal approach, the dilution effect – assuming it is effective – should be stronger in Italy than in France.

(2) *The dilution effect should be weaker in the most affluent families*. Educating children involves considerable fixed direct and indirect costs. From a relative viewpoint, these costs are higher for poorer families.

(3) *The dilution effect should weaken for the youngest cohorts, mainly for obtaining a secondary degree*, as, over the last century, Italy and France have experienced a progressive expansion of participation in the school system. Consequently, a growing proportion of people have obtained a secondary school or a university qualification. Today, it has become "normal" for young people to earn a secondary degree, regardless of gender, family social class, and – presumably – family size.

(4) *There could be differential advantages and drawbacks based on birth order in pursuing education.* As previously stated, the literature provides mixed results regarding birth order, and does not seem to support the assumption that a specific position confers clear advantages. However, in this research project, the birth order is assumed to act negatively. In other words, following Zarca's findings on intergenerational mobility, it is assumed that the first-born child is more likely to achieve top educational levels than the second-born child, who, in turn, has greater opportunities than the third-born child, and so on (Zarca 1995a, 1995b).

2.3 Data

Data for Italy come from the *Indagine Multiscopo sulle famiglie: Famiglia e soggetti sociali (2003)* (The Multipurpose Survey on the Family), a representative national survey conducted by the Italian National Statistics Institute (Istat) on a sample of about 24,000 Italian families and 50,000 individuals. Data for France have been taken from the *Etude de l'Histoire Familiale (1999)* (the Family History Survey), a representative national survey conducted by the French National Institute of Statistics and Economic Studies (Insee) and the French National Institute for Demographic Studies (INED) on a sample of 380,000 respondents. Despite the differences between the two surveys, a close correspondence has been maintained, allowing for strong comparability in terms of the topics and the time periods observed. The response rate was 18% for the Italian survey (personal communication with Lidia Gargiulo of Istat, E-mail: gargiulo@istat.it, see also Brancato et al. 2004) and 21% for the French survey (Barre and Vanderschelden 2004, page 7).

Additionally, both surveys include detailed information on *individual education* (i.e., the response variable), and on the *number* and *birth order* of siblings (i.e., the main covariates). The data also allowed us to create homogeneous categories concerning the *socioeconomic backgrounds* of the family, based on the parents' occupations and the mother's activity status.

3. Descriptive analysis: Method and variables

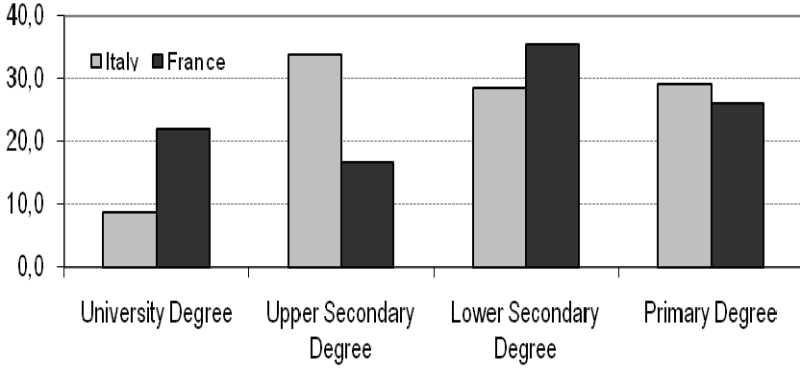
3.1 Method

Several logistic regression models have been employed to describe the impact of the number of siblings and birth order on the probability of achieving at least a secondary school (or university) qualification, controlling for other covariates. The dummy response variable juxtaposes high and low educational levels; "success" is defined as having completed secondary school or university.

3.2 Response variable

The response variable is *educational level,* which is available from the Italian survey as the last qualification achieved, rather than total years of education. For this reason, the variable has been divided into four categories: (1) university degree, (2) upper secondary school diploma, (3) lower secondary school diploma, and (4) primary school diploma. Despite the differences between the French and Italian educational systems, this classification is closely tied to country-specific levels. When the analysis is concerned exclusively with the probability of reaching a university degree, only individuals aged 25 or older are considered. By contrast, when modeling the probability of achieving an upper secondary diploma, the whole sample of individuals aged 18 and older are included. Since the question concerning educational qualification is asked in the first and more general part of the questionnaire, and is compulsory, the participation rate is 100% for the relative age group. Figure 1 shows the distribution of educational qualifications when comparing the two countries. In France, the proportion of people with university degrees is double that of Italy, while in Italy, the proportion of people with upper secondary degrees is twice that of France.

Figure 1: Distribution by education in Italy and France. People aged 18+ interviewed in 2003 (Italy) and 1999 (France)



3.3 Covariates

The main explanatory variable is the total number of brothers and sisters. The distribution of large and small families has changed considerably over the last century, towards a greater number of smaller families in the youngest cohort (table 1). The latter trend implies the "mandatory" inclusion of age classes when investigating the effect of number of siblings on education.

Table 1: Frequency distribution of family sizes according to age, Italy and France (%)

Ages *(cohorts)*	Mean (all ages)		18-24 *(ITA:1979-1985) (FRA: 1975-1981)*		25-44 *(ITA:1959-1978) (FRA: 1955-1974)*		45-64 *(ITA:1939-1958) (FRA: 1935-1954)*		65 + *(ITA:1901-1938) (FRA: 1900-1934)*	
N. of siblings	Italy	France	Italy	France	Italy	France	Italy	France	Italy	France
Only Child	11.7	11.1	13.3	10.8	11.9	9.0	12.0	10.7	10.3	15.9
1 sibling	29.6	22.8	51.2	35.9	36.9	24.1	24.4	18.6	15.6	20.5
2 siblings	21.8	21.6	25.1	28.4	24.9	23.5	20.6	19.4	17.1	18.2
3 siblings	13.3	14.3	7.3	11.7	12.1	14.8	15.2	15.0	15.2	13.4
4 siblings	8.0	9.5	1.8	5.6	6.0	9.3	9.1	10.9	12.5	9.7
5 +	15.6	20.7	1.4	7.7	8.3	19.4	18.7	25.4	29.3	22.3
Total	100	100	100	100	100	100	100	100	100	100
N	*41 559*	*366 231*	*4 378*	*34 898*	*15 280*	*140 798*	*12 529*	*113 393*	*9 372*	*77 142*

Parents' occupation has been used as a proxy of socioeconomic conditions. The variable was created by combining the father's and the mother's jobs, with the most well-paid of the two used as an indicator of the general family situation. Occupations were grouped into four categories: (1) higher level occupations, (2) white collar, (3) self-employed, and (4) blue collar and unemployed. The decision to analyze the self-employed category separately is based on the widely observed phenomenon of intergenerational transmission in that particular category: i.e., there are reasons to believe that the sons of the self-employed are likely to follow in their parents' footsteps (Zarca 1995a). We assume that the occupational status of a parent could indirectly affect the children's educational careers, encouraging them, for example, to leave school earlier to join the family business, or to remain in the school system in order to acquire the knowledge required for a specific occupation. Table 2 shows the distribution by education for each parent's socioeconomic category. In both countries, about two-thirds of the children of parents in the higher level occupations achieve at least a secondary school diploma. In Italy the same pattern holds for those whose parents are white collar workers, while in France only about the half of the children whose parents belong to this category earn at least a secondary school diploma. Generally speaking, the association between family social class and education is stronger in Italy than in France.

Table 2: Frequency distribution of levels of education according to socioeconomic origin, Italy and France (%)

Parents' Job	Higher Level		White Collar		Self-Employed		Blue Collar	
Education	Italy	France	Italy	France	Italy	France	Italy	France
University Degree	27.6	52.7	21.1	27.4	6.2	19.6	3.8	8.8
Upper Secondary Diploma	44.8	21.1	52.3	20.1	27.4	19.2	25.0	11.0
Lower Secondary Diploma	16.8	20.7	18.9	38.0	26.5	36.9	32.5	39.1
Primary Diploma	10.8	5.5	7.7	14.5	39.9	24.3	38.7	41.1
Total	100	100	100	100	100	100	100	100

The models also include a covariate expressing the *mother's activity status*, in order to better characterize the family's socioeconomic background, as well as the presence of parents at home. Indeed, a study on the effects of the number of siblings, birth order, and social origins on children's educational outcomes shows that having a retired father increases the children's educational success (Gary-Bobo et al. 2006). This is probably linked to the additional support available for children at home, which in the present analysis is assumed to be provided by the mother.

The *birth order* is another important variable related to the sibship composition, and it is strictly connected to the number of siblings. As we noted in the literature review, the role of birth order on child outcomes remains controversial, and it is thus worth taking into account. In this study, birth order will be included in interaction with family size, in order to investigate how children's opportunities change based on whether they belong to different sub-groups.

The demographic characteristics *age* and *gender* are also included. Moreover, the year at birth has been centered to its median value to obtain more reliable estimates.

Tables 3 and 4 complete the data description. The frequency distribution of response variable has been summarized for particular sub-populations, and is split by age, gender, and family size (two or fewer siblings, and three or more siblings). The data displayed in these tables appear to support the dilution effect, as the proportion of people with university degrees or upper secondary diplomas is shown to be consistently smaller in larger families.

Table 3: Some socioeconomic characteristics according to age classes, sex, and family size. Italy, column %

Italy %	25-44 years old				45-64 years old				65+ years old			
	Males		Females		Males		Females		Males		Females	
Siblings	2 or less	3 or more	2 or less	3 or more	2 or less	3 or more	2 or less	3 or more	2 or less	3 or more	2 or less	3 or more
Education												
University Degree	13.1	6.3	16.6	7.6	12.8	5.6	11.7	4.3	7.5	3.9	3.3	1.3
Upper Secondary Diploma	50.2	31.3	53.6	33.0	36.1	19.8	32.1	17.0	18.3	9.1	12.9	5.6
Lower Secondary Diploma	34.1	51.4	27.3	48.1	31.9	33.7	28.0	26.7	19.7	12.1	14.2	7.6
Primary Diploma	2.6	11.1	2.5	11.4	19.2	40.9	28.1	52.0	54.5	74.9	69.7	85.6
Parents' Occupation												
Higher Level Occupation	15.2	9.6	13.9	8.7	9.8	7.1	8.7	6.4	7.2	5.9	7.6	5.1
White Collar	21.0	11.8	23.4	13.2	13.6	7.9	15.8	8.3	9.7	4.6	8.9	5.6
Self-Employed	20.2	22.8	20.7	21.3	24.1	28.0	23.8	25.3	28.6	35.2	28.4	31.3
Blue Collar	43.6	55.8	42.1	56.8	52.6	75.1	51.8	60.0	54.5	54.3	55.1	58.0
Mother's Activity Status												
Inactive	59.0	74.9	57.8	64.1	67.5	75.2	64.1	74.1	66.3	69.8	67.8	69.2
Active	41.0	25.1	42.2	35.9	32.5	24.8	35.9	25.9	33.7	30.3	32.2	30.8

Table 4: Some socioeconomic characteristics according to age classes, sex and family size. France, column %

France %	25-44 years old				45-64 years old				65+ years old			
	Males		Females		Males		Females		Males		Females	
Siblings	2 or less	3 or more	2 or less	3 or more	2 or less	3 or more	2 or less	3 or more	2 or less	3 or more	2 or less	3 or more
Education												
University Degree	36.7	19.3	42.5	19.6	23.5	13.8	26.2	10.4	12.4	6.5	4.9	2.5
Upper Secondary Diploma	18.6	13.9	21.9	18.6	18.0	11.4	34.1	13.6	11.7	6.8	12.0	6.7
Lower Secondary Diploma	41.1	57.8	32.2	51.2	35.1	36.6	21.3	32.2	24.7	18.6	26.6	17.4
Primary Diploma	3.5	9.1	3.3	10.6	23.5	38.1	18.4	44.0	51.3	68.1	56.5	73.4
Parents' Occupation												
Higher Level Occupation	18.1	9.0	18.7	9.0	11.7	7.8	11.8	7.5	7.9	4.7	7.0	4.4
White Collar	41.8	27.7	42.0	28.4	29.0	21.4	29.4	21.7	22.6	14.3	21.1	15.2
Self-Employed	9.0	8.0	9.6	8.0	12.6	9.4	12.9	9.2	14.9	12.2	15.5	11.6
Blue Collar	31.2	55.4	29.7	54.6	46.7	61.3	45.9	61.5	54.7	68.9	55.4	68.8
Mother's Activity Status												
Inactive	36.2	60.3	35.2	59.9	52.7	65.7	51.1	64.8	55.7	61.7	55.2	60.9
Active	63.8	39.7	64.8	40.2	47.3	34.3	48.9	35.2	44.3	38.3	44.8	39.1

4. Descriptive analysis: Results

This section presents the results of the logistic regression models, which test the four hypotheses outlined in Section 2. First, the relative risks of achieving university and secondary school qualifications will be presented without including any interaction, comparing France and Italy (Model 1). Second, the probability of reaching the two possible educational levels will be analyzed, including interactions with the parents' occupation categories (Model 2) and age classes (Model 3). Finally, an interaction between the family size and birth order will be included in the basic model (Model 4).

4.1 Testing the first hypothesis: The dilution effect holds in both countries, but is stronger in Italy

Table 5 displays the probabilities of achieving university and secondary school qualifications according to different family sizes, controlling for parents' occupation, the mother's activity status, sex, and age (Model 1). In both Italy and France, there are no significant differences between only children and people with one sibling. By

contrast, people who have two siblings or more are clearly penalized. Despite our initial assumptions, family size is shown to have a strong negative effect in France, although this effect is found to be weaker than in Italy. People who are from a disadvantaged socioeconomic background are shown to have lower chances of achieving high levels of education in France and in Italy. Furthermore, in line with our initial assumptions, the findings suggest that the mother's activity status has a negative influence on the children's education. Finally, our results indicate that sex differences, where they exist, are not very important in determining the opportunities of individuals.

Table 5: Effects of family size and other covariates on the probability of attaining university and secondary school degrees

	University degree				Secondary diploma			
	Italy exp(β)		France exp(β)		Italy exp(β)		France exp(β)	
Number of siblings								
0	1.00	n.s.	0.95	**	0.99	n.s.	0.97	*
1 (Ref)	1		1		1		1	
2	0.75	***	0.82	***	0.67	***	0.78	***
3	0.62	***	0.64	***	0.50	***	0.61	***
4	0.42	***	0.53	***	0.33	***	0.50	***
5 and more	0.29	***	0.38	***	0.22	***	0.36	***
Parents' Occupation								
Higher Level	9.34	***	9.00	***	5.92	***	9.17	***
White –Collar	5.59	***	2.87	***	4.63	***	2.78	***
Self-Employed	2.00	***	2.55	***	1.52	***	2.55	***
Blue Collar (Ref)	1		1		1		1	
Mother's Activity Status								
Active	0.90	**	0.89	***	0.84	n.s.	0.83	***
Sex								
Males	1.00	n.s.	0.91	***	0.99	***	1.13	***
Age	0.99	***	0.96	***	0.96	***	0.96	***

n.s. p>0.10; * 0.05<p<0.10; ** 0.01<p<0.05; *** p<0.01

4.2 Testing the second hypothesis: The dilution effect is weaker among French bourgeois families

Tables 6 and 7 show estimates obtained from Model 2. The probability of earning university and secondary school qualifications decreases significantly as the family size increases, but the decrease is slower in France for almost all the socioeconomic groups. Nonetheless, there is an interesting result among people who belong to large families in the top socioeconomic group: in France, among wealthier children who have five siblings or more, the probability of earning a university degree is reduced by "only" 40%, and the chances of earning a high school diploma are lowered by just one-half, with respect to the reference group. By contrast, in all the other categories in France, and also in Italy, the probability of achieving top qualifications is about 80% lower. Moreover, there are no significant differences between only children and people with one sibling, and this holds true for both countries, with the following exceptions: (1) French only children whose parents are in the blue collar category seem to be less likely to reach top levels than the reference group, and (2) French children whose parents are self-employed turn out to be more advantaged than the reference category.

Table 6: Effect of family size on educational outcomes by socioeconomic origin, Italy and France, university degrees

	University Degree											
	Higher Level		White Collar		Self-Employed				Blue Collar			
	Italy exp(β)	France exp(β)	Italy exp(β)	France exp(β)	Italy exp(β)		France exp(β)		Italy exp(β)		France exp(β)	
Number of siblings												
0	1.01 n.s.	0.92 *	1.17 n.s.	0.97 n.s	0.97	n.s.	0.97	n.s.	0.87	n.s.	0.83	***
1 (Ref)	1	1	1	1	1		1		1		1	
2	0.80 *	0.97 n.s.	0.82 *	0.75 ***	0.77	**	0.87	***	0.70	***	0.80	***
3	0.70 **	0.89 ***	0.83 n.s.	0.55 ***	0.62	***	0.73	***	0.44	***	0.62	***
4	0.50 ***	0.83 ***	0.50 ***	0.44 ***	0.36	***	0.57	***	0.39	***	0.50	***
5 and more	0.22 ***	0.59 ***	0.30 ***	0.33 ***	0.27	***	0.47	***	0.32	***	0.36	***
Sex												
Males	1.00 ***	0.92 ***	1.10 ***	1.07 ***	1.05	***	0.79	***	0.86	***	1.00	***
Age	1.01 n.s.	0.98 ***	1.01 n.s.	0.97 ***	0.99	n.s.	0.97	***	0.99	n.s.	0.95	n.s.

n.s. p>0.10; * 0.05<p<0.10; ** 0.01<p<0.05; *** p<0.01

Table 7: Effect of family size on educational outcomes by socioeconomic origin, Italy and France, secondary school diploma

	Secondary School Diploma															
	Higher Level				White Collar				Self-Employed				Blue Collar			
	Italy exp(β)		France exp(β)		Italy exp(β)		France exp(β)		Italy exp(β)		France exp(β)		Italy exp(β)		France exp(β)	
Number of siblings																
0	0.95	n.s.	0.96	n.s.	1.03	n.s.	1.02	n.s	1.01	n.s.	1.10	**	0.98	n.s.	0.89	***
1 (Ref)	1		1		1		1		1		1		1		1	
2	0.73	***	0.86	***	0.76	***	0.74	***	0.70	***	0.89	***	0.63	***	0.78	***
3	0.55	***	0.77	***	0.60	***	0.54	***	0.53	***	0.74	***	0.46	***	0.62	***
4	0.31	***	0.74	***	0.36	***	0.45	***	0.38	***	0.54	***	0.32	***	0.50	***
5 and more	0.19	***	0.50	***	0.18	***	0.32	***	0.25	***	0.44	***	0.23	***	0.37	***
Sex																
Males	1.02	***	1.01	***	1.23	***	1.18	***	1.09	***	0.99	***	0.87	***	1.17	***
Age	0.98	n.s.	0.98	n.s.	0.97	***	0.97	***	0.95	*	0.97	n.s.	0.96	***	0.96	***

n.s. p>0.10; * 0.05<p<0.10; ** 0.01<p<0.05; *** p<0.01

4.3 Testing the third hypothesis: No differences by cohort

In Tables 8 and 9, the dilution effect by age is presented. In both countries, the impact of belonging to a large family is found to be negative throughout the 20th century, despite tremendous changes in the educational systems and the diffusion of secondary education. This remains true for every level of education analyzed, even when we control for the parents' occupation, the mother's activity status, sex, and year of birth. As previously noted, however, the dilution effect is stronger in Italy than in France; this holds true for each cohort group considered. This stability is especially striking in light of other fundamental changes. For example, in cohort after cohort, the relative impact of the parents' occupation becomes less and less important in determining children's educational outcomes. Moreover, the impact of the mother's occupation and sex have reversed over time: i.e., girls with working mothers born in the second half of the 20th century succeeded more frequently in school.

Table 8: Effect of family size on educational outcomes by age classes, Italy and France, university degree

	University Degree					
Age Class *(COHORT)*	25-44 *(ITA:1959-1978) (FRA: 1955-1974)*		45-64 *(ITA:1939-1958) (FRA: 1935-1954)*		65+ *(ITA:1901-1938) (FRA: 1900-1934)*	
	Italy exp(β)	France exp(β)	Italy exp(β)	France exp(β)	Italy exp(β)	France exp(β)
Number of Siblings						
0	0.95 n.s.	0.94 **	1.10 n.s.	0.98 n.s.	0.95 n.s.	1.08 n.s.
1 (Ref)	1	1	1	1	1	1
2	0.74 ***	0.79 ***	0.75 ***	0.87 ***	0.73 n.s.	0.87 ***
3	0.61 ***	0.60 ***	0.55 ***	0.71 ***	0.67 ***	0.78 ***
4	0.51 ***	0.46 ***	0.33 ***	0.63 ***	0.35 ***	0.65 ***
5 and more	0.20 ***	0.35 ***	0.30 ***	0.44 ***	0.37 ***	0.42 ***
Father's Occupation						
Higher Level	8.20 ***	6.77 ***	10.27 ***	11.29 ***	15.88 ***	19.27 ***
White Collar	5.04 ***	2.38 ***	6.51 ***	3.28 ***	6.52 ***	4.75 ***
Self-Employed	2.16 ***	2.10 ***	1.93 ***	3.06 ***	1.77 **	3.91 ***
Blue Collar (Ref)	1	1	1	1	1	1
Mother's Activity Status						
Active	1.09 n.s.	0.96 **	0.76 ***	0.90 ***	0.69 **	0.64 ***
Sex						
Males	1.33 ***	1.20 ***	0.84 ***	0.70 ***	0.35 ***	0.34 ***
Age	1.02 ***	0.95 ***	0.96 ***	0.95 ***	0.98 *	0.99 **

n.s. p>0.10; * 0.05<p<0.10; ** 0.01<p<0.05; *** p<0.01

Table 9: Effect of family size on educational outcomes by cohorts, Italy and France, secondary school diploma

	Secondary School Diploma							
Age Class *(COHORT)*	25-44 *(ITA:1959-1978)* *(FRA: 1955-1974)*		45-64 *(ITA:1939-1958)* *(FRA: 1935-1954)*		65+ *(ITA:1901-1938)* *(FRA: 1900-1934)*	
	Italy exp(β)	France exp(β)	Italy exp(β)	France exp(β)	Italy exp(β)	France exp(β)
Number of Siblings						
0	0.88 *	0.92 **	1.17 **	1.01 n.s.	1.07 n.s.	1.16 ***
1 (Ref)	1	1	1	1	1	1
2	0.65 ***	0.76 ***	0.68 ***	0.81 ***	0.64 ***	0.83 ***
3	0.48 ***	0.57 ***	0.47 ***	0.65 ***	0.59 ***	0.71 ***
4	0.30 ***	0.44 ***	0.34 ***	0.54 ***	0.33 ***	0.61 ***
5 and more	0.17 ***	0.33 ***	0.22 ***	0.37 ***	0.27 ***	0.42 ***
Father's Occupation						
Higher Level	5.03 ***	6.93 ***	7.26 ***	10.76 ***	12.44 ***	16.94 ***
White Collar	4.27 ***	2.32 ***	6.35 ***	3.11 ***	6.76 ***	4.19 ***
Self-Employed	1.55 ***	2.16 ***	1.73 ***	2.96 ***	1.45 ***	3.61 ***
Blue –Collar (Ref)	1	1	1	1	1	1
Mother's Activity Status						
Active	1.03 n.s.	0.96 **	0.75 ***	0.85 ***	0.53 ***	0.54 ***
Sex						
Males	1.33 ***	1.41 ***	0.77 ***	0.92 ***	0.49 ***	0.63 ***
Age	0.96 ***	0.96 ***	0.92 ***	0.95 ***	0.96 ***	0.99 ***

n.s. p>0.10; * 0.05<p<0.10; ** 0.01<p<0.05; *** p<0.01

4.4 Testing the fourth hypothesis: It is better to be the first-born child

Our last model includes the interaction between the number of siblings and birth order. Although the primary focus of our analysis is on the impact of family size, and not of birth order, this allows us to observe whether there are differences in educational outcomes according to birth position within families of the same size. Firstborns in two-sibling families are used as the reference category. Except for only children and the second-born of two-sibling families, all the positions are significantly disadvantaged in terms of educational achievement. A further check on possible birth order disparities within each family size was also performed, again comparing firstborns (in families of three and four siblings) with the other positions (see the asterisks in brackets in Table 10). We tested whether differences between pairs of coefficients were significant, keeping firstborns as the reference category. In Italy, within families of three or four

children, we found that only the lastborn is significantly disadvantaged in terms of education; in fact, our analysis shows that the middle-born in families of three and four children has a probability of obtaining a level of education not significantly different from that of the firstborn. The only exception concerns the probability that the lastborn within a three-sibling family will earn a university degree, which is found to be not significantly different from that of his or her older peers. Moreover, in France, the effect of birth order is found to be strong and regular, regardless of family size, for both levels of education: i.e., the firstborn is favored, the lastborn is penalized, and those in the middle are in an intermediate position.

5. Toward a causal approach

The models identified so far do not resolve the issue of causality. Indeed, the observed dilution effect could be biased due to unobserved heterogeneity. In particular, parental preferences regarding their children's education could affect both the number of children they decide to have, and their children's future education. In other words, couples who have a strong desire for "high quality" children would try to keep the family size small in order to ensure that each child receives the desired amount of material and non-material resources. By contrast, those parents who choose to have many children might be characterized as having less interest in their children's education. Because a parental heterogeneous preference may exist, additional "adjustments" are needed in order to estimate correctly the effect of family size. But parental attitudes regarding their children's educational future can be only one of several possible sources of heterogeneity: indeed, many other factors (e.g., where the family lives and whether they are recent immigrants) might also influence both the total number of children and their educational careers.

In this study, this issue is resolved through the use of an instrumental variable which exogenously accounts for family size variation. This variable respects two conditions: (1) it is directly linked to the number of siblings (the main explanatory variable), and (2) it is connected to educational level (the response variable) only by means of the main explanatory variable. A common variable used in the literature is the twin variable (Black, Devereux, and Salvanes 2005; Angrist, Lavy, and Schlosser 2006). Other authors (Merlier and Monso 2007; Maurin and Moschion 2006) also use the age distance between parents or the TFR of the mother's birth cohort. Unfortunately, the only instrumental variable available from our data is the **sex of the two eldest children** (same vs. different). However, this variable can only be calculated for families with two or three children (see Appendix for a detailed construction).

Table 10: Odds Ratios of achieving high levels of education, model with interaction between family size and birth order ([a])

	University Degree				Secondary School Diploma			
	Italy exp(β)		France exp(β)		Italy exp(β)		France exp(β)	
Siblings*Birth Order								
Only Child	0.96	n.s.	0.97	n.s.	0.94	n.s.	0.95	*
1st between 2 (Ref)	1		1		1		1	
2nd of 2	0.90	n.s.	0.85	***	0.89	n.s.	0.84	***
1st of 3	0.86	*	0.85	***	0.70	***	0.81	***
(Ref)		-		-		-		-
2nd of 3	0.63	***	0.71	***	0.58	***	0.67	***
		(#)		*(**)*		*(#)*		*(***)*
3rd of 3	0.69	***	0.69	***	0.68	***	0.64	***
		(#)		*(**)*		*(*)*		*(***)*
1st of 4	0.81	n.s.	0.70	***	0.55	***	0.67	***
(Ref)		-		-		-		-
2nd of 4	0.53	***	0.58	***	0.44	***	0.52	***
		(#)		*(**)*		*(#)*		*(***)*
3rd of 4	0.54	***	0.55	***	0.47	***	0.51	***
		(#)		*(*)*		*(#)*		*(***)*
4th of 4	0.51	***	0.53	***	0.47	***	0.49	***
		(*)		*(**)*		*(*)*		*(***)*
Father's Occupation								
Higher Level	9.38	***	8.09	***	5.91	***	8.46	***
White Collar	5.58	***	2.80	***	4.64	***	2.77	***
Self-Employed	2.06	***	2.43	***	1.50	***	2.47	***
Blue Collar (Ref)	1		1		1		1	
Mother's Activity Status								
Active	0.93	n.s.	0.87	***	0.87	***	0.80	***
Sex *Males*	1.04	n.s.	0.96	***	1.03	n.s.	1.18	***
Age	0.99	***	0.96	***	0.96	***	0.96	***

n.s. / # p>0.10; * 0.05<p<0.10; ** 0.01<p<0.05; *** p<0.01

([a]) *In brackets the statistical significance of comparison with (Ref.) of the same sibling size.*

The data for both countries show that couples with two same-sex children (either two boys or two girls) are more likely to have a third child than those who already have two opposite-sex children. Specifically, the proportion of the former exceeds that of the latter by 7.8 percentage points for Italy, and by 8.2 percentage points for France (Table 11). Consequently, the first condition for being an instrumental variable is respected. The second condition also holds, as the parental preference for having children with different genders does not itself have any impact on the children's educational attainment; its impact is solely related to the effect of increasing the family size.

Table 11: Proportion of families progressing from two to three offspring, according to the sex of the two oldest children

	Italy (%)	France (%)
(a) Two boys	41.4	54.3
(b) Two girls	43.2	49.7
Same-sex (average of a & b)	42.3	52.0
Mixed sex	34.5	43.8

It may be useful to explain in more detail the meaning of this statistical device for assessing the causal link between variable, following the approach of Agrist et al. (1996). If, for each individual *i*, Y_i is the observed educational outcome, D_i is the observed "treatment" (i.e., belonging to a two- or three-child family), and Z_i is the instrument defined above, the model specification is the following:

$$Y_i = \beta_0 + \beta_1 \cdot D_i + \varepsilon_i \quad (1)$$

$$D_i^* = \alpha_0 + \alpha_1 \cdot Z_i + \upsilon_i \quad (2)$$

$$D_i^* = \begin{cases} 1 & if \quad D_i^* > 0 \\ 0 & if \quad D_i^* \le 0 \end{cases} \quad (3)$$

In this model, β_1 (i.e., the parameter relative to the treatment) is the causal effect of D on Y. In order to **correctly** identify β_1, and thus to obtain an unbiased estimation of the causal effect, two assumptions must hold true. (1) *The covariance between the instrumental variable and the main covariate must differ from zero*. In our analysis, this is straightforward: the parity progression to a third child obviously increases the family

size. This condition is verified empirically by looking at the coefficient of the first stage regression: for both Italian and French data, α_1 turns out to be different from zero. (2) *Any effect of Z on Y must be through an effect of Z on D.* This means that, together with the absence of Z in Equation (1), Z_i is uncorrelated with error terms ε_i and υ_i (orthogonal error process):

$$E[Z_i \cdot \varepsilon_i] = 0, \qquad E[Z_i \cdot \upsilon_i] = 0$$

This instrument exploits the widely observed phenomenon of parental preferences for a mixed sibling sex composition. In particular, parents of same-sex siblings are significantly more likely to go on to have an additional child. Because the gender mix is almost randomly assigned, a dummy for whether the sex of the first two children matches provides a plausible instrument for representing further childbearing among women with at least two children (Angrist and Evans 1998). If the coefficients related to the family size change to positive after the instrument is included in the model, this means that the educational advantage of being reared in a small family is not related to the family size in itself, but rather to some other unobserved characteristics which distinguish these kinds of children from those who were reared in a bigger family.

The instrument was created as an indicator of whether the first two children of two- or three-child families were of the same sex, and is applied to the variable that discriminates between families with two or three children. Table 12 compares the logit model and the 2SLS estimation for the probability of earning a university degree in Italy or a secondary school diploma in France. The parameter estimates of the "true" effect of progressing from two to three children on the probability of obtaining a university or secondary school qualification turn out to be insignificant in the Italian case, although they remain negative. By contrast, in the French context, an increase in family size due to the sex composition of siblings *raises* the probability that the children will attain the analyzed levels of education. When the instrumental variable is introduced, the transition from two to three children *increases* (rather than decreases) the probability that each of the siblings will earn a university degree by 13 percentage points, and a secondary school diploma by 22 percentage points. The new "true" effect is "depurated" from the sources of heterogeneity linked to parental preferences towards their children's education. Although the meaning of these results should be further investigated, the differences between Italy and France may be due to the existence of better policies in France aimed at families with three children. After controlling for the potential endogeneity of the quality vs. quantity choice, we find that French children in large families are favored, whereas in Italy they are penalized.

Table 12: The effect of family size on obtaining university or secondary qualifications. Logit regression and 2SLS regression (causal analysis with Instrumental Variable IV).

	A. with covariates							
	Italy				France			
Dependent variable	logit		IV		logit		IV	
University Degree	-0.30	***	-0.21	n.s.	-0.21	***	0.13	**
s.e.	*(0.0597)*		*(0.2631)*		*(0.0158)*		*(0.0543)*	
Secondary Diploma	-0.40	***	0.30	n.s.	-0.26	***	0.22	***
s.e.	*(0.0376)*		*(0.3750)*		*(0.0136)*		*(0.0567)*	
	B. without covariates							
University Degree	-0.44	***	-0.30	n.s.	-0.25	***	0.27	***
s.e.	*(0.0570)*		*(0.2529)*		*(0.0142)*		*(0.0637)*	
Secondary Diploma	-0.56	***	0.14	n.s.	-0.29	***	0.38	***
s.e.	*(0.0338)*		*(0.3841)*		*(0.0119)*		*(0.0684)*	

Our decision to employ an instrumental variable of this kind can be critiqued. Furthermore, as we will discuss below, the issues explored in this study should be investigated in greater depth. For example, the sex ratio may vary systematically with maternal and/or paternal age. Moreover, the connection between the IV and the main explanatory variable, while significant, is not very strong. Finally, we consider only families with two or three siblings, which make up only part of the sample, especially in France. These aspects leave room for further research, as we will describe below.

6. Discussion and conclusion

The descriptive analysis shows that, mainly in Italy, but also in France, being an only child or having only one sibling is associated with a greater probability of obtaining at least a high school qualification, even after controlling for the parents' job, the mother's working status, sex, and the year of birth. Our first hypothesis is confirmed (see Part 2). Moreover, the second hypothesis is partially confirmed: the dilution effect is found to be weaker for wealthy people, mainly in France. Meanwhile, the third hypothesis is not proved by the data, because the interaction with cohorts results in an unchanged pattern over the course of the 20th century. Finally, the birth order is shown to influence the probability of achieving a higher level of education: in France, the effect of birth order is found to be stronger than in Italy, with the lastborn being more strongly penalized.

The main result of the descriptive analysis is important, as it shows that the probability of accumulating rich human capital correlates to family size, even after a number of covariates have been taken into account (mainly the social class and birth cohorts of the parents). Indeed, when the sex of the two eldest children (same vs. different) is used as an instrumental variable for family size in order to avoid the endogeneity associated with the latter, the effect of family size (two vs. three children) on education, while negative, is not found to be statistically significant in Italy; whereas in France, it is shown to become positive. Overall, our causal analysis is far from satisfactory. It examines only families with two or three children, and the instrumental variable is weak; i.e., its correlation with family size is moderate. While our results are in line with the findings of certain studies that use stronger instrumental variables in examining France (Merllié and Monso 2007) and a number of other countries (Angrist, Lavy, and Schlosser 2005); in other contexts, the effect of sib-size on children education has been shown to be "purely casual" (Jæger 2008). Thus, this avenue of research should be further pursued for both France and, especially, Italy, possibly by looking at other datasets, and/or by trying to find better instruments or other techniques for teasing out the "true" causality of these trends.

Although this final part of our analysis clearly leaves room for additional research, we believe that our results can provide insight into this issue for those who are seeking to increase the human capital of young people through fiscal and/or social policies. Beyond the causal links which generate our results, we can assert that young people with two or more siblings appear to be strongly penalized in terms of scholastic opportunities. Given this clear evidence that there is no difference—in either France or Italy—in the probability that only children and those who have only one sibling will achieve higher levels of education, *a greater share of resources should be allocated to helping young people with two or more siblings acquire more human capital, rather than provided to all children equally, as young people from larger families may need*

additional assistance in overcoming both the objective economic problems faced by their families, and the cultural opposition to education of their parents.

7. Acknowledgements

Special thanks go to Laurent Toulemon, without whose contributions this project would not have been possible. We also wish to thank Enrico Rettore, Luca Pieroni, and Francesco Billari for their useful suggestions on econometrics; and Olivier Monso, Ugo Trivellato, and Guglielmo Weber for their valuable comments.

References

Angrist, J.D. and Evans, W.N. (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *The American Economic Review* 88(3): 450-477.

Angrist, J.D., Imbens, G.W., and Rubin, D.B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91(434): 444-472. doi:10.2307/2291629.

Angrist, J.D., Lavy, V., and Schlosser, A. (2005). New evidence on the causal link between quantity and quality of children. Boston, MA, USA (MIT working paper).

Barre, C. and Vanderschelden, M. (2004). L'enquête «Étude de l'histoire familiale» de 1999: Résultats détaillés. *Insee Résultats Société* 33 August.

Black, S.E., Devereux, P.J., and Salvanes, K.G. (2005). The more the merrier? The effect of family size and birth order on children's education. *Quarterly Journal of Economics* 120(2): 669-700. doi:10.1162/0033553053970179.

Blake, J. (1980). *Family Size and Achievement*. Berkeley: University of California Press. http://ark.cdlib.org/ark:/13030/ft6489p0rr/.

Brancato, G., Pellegrini, C., Simeoni, G., and Signore, M. (2004). *Standardising, evaluating and documenting quality: The implementation of Istat information system for survey documentation – SIDI*. Proceedings of the European Conference on Quality and Methodology in Official Statistics (Q2004). Mainz, May, 24-26.

Conley, D. and Glauber, R. (2005). Parental education investment and children's academic risk: Estimates of the impact of sibship size and birth order from exogenous variation in fertility. (NBER working paper 11302).

Dalla Zuanna, G. (2007). Social mobility and fertility. *Demographic Research* 17(15): 441-464. doi:10.4054/DemRes.2007.17.15.

French National Statistical Institute (Insee). (2000). Courrier des statistiques. (93, March).

Gary-Bobo, R.J., Prieto, A. and Picard, N. (2006). Birth-order and sibship sex-composition effects in the study of education and earnings. London, UK. (CEPR Discussion paper 5514)

Italian National Statistical Institute (Istat). (2003). Famiglia. soggetti sociali e condizione dell'infanzia. La mobilità sociale. indagine multiscopo sulle famiglie, 1998. Collana Informazioni: Roma, Italy.

Jæger, M.M. (2008). Do large sibships really lead to lower educational attainment? New evidence from quasi-experimental variation in couples' reproductive capacity. *Acta Sociologica* 51(3): 217-235. doi:10.1177/0001699308094167.

Maurin, E. and Moschion, J. (2006). The social multiplier and labour market participation of mothers. Bonn, Germany. (IZA Discussion Paper 2513).

Merllié, D. and Monso, O. (2007). La destinée sociale varie avec le nombre de frères et sœurs, INSEE, France, portrait social.

Steelman, L.C., Powell, B., Werum, R., and Carter, S. (2002). Reconsidering the effects of siblings configuration: Recent advances and challenges. *Annual Review of Sociology* 28(1): 243-269. doi:10.1146/annurev.soc.28.111301.093304.

Zarca, B. (1995a). L'héritage et la mobilité sociale au sein de la fratrie. I. L'héritage et la mobilité sociale différentielle des frères. *Population (French Edition)* 50(2): 331-356. doi:10.2307/1534192.

Zarca, B. (1995b). L'héritage et la mobilité sociale au sein de la fratrie. II. L'héritage et la mobilité sociale différentielle des sœurs. *Population (French Edition)* 50(4/5): 1137-1154. doi:10.2307/1534318.

Appendix: Constructing the instrumental variable

The instrumental variable "same sex of the two eldest children" has been created as an indicator of whether the first two children of a family of two or three children were of the same sex.

The construction of the instrument is quite straightforward for families composed of two children. Indeed, there are eight (=2x2²) possible combinations of the compositions of siblings by sex and birth order; four for each birth order considered (Table A1). Specifically, the information on siblings is collected at the individual level, and, for each person, data are available on the number of brothers and sisters, as well as on the birth order. Moreover, since the individual who was interviewed can either be the first- or the second-born, two different cases are possible for each composition. When comparing, for example, Case 3_1st with Case 4_2nd, we find that the group is always composed of a girl and a boy; but that, in Case 3, the information is available for the first-born girl, while in Case 4, the information is for the second-born boy.

Table A1: Possible sibling sex compositions and instrumental variable value among two-child families[3]

Case	Firstborns (1st)	Z	Second-borns (2nd)	Z
1	GG	1	GG	1
2	BB	1	BB	1
3	GB	0	BG	0
4	BG	0	GB	0

Concerning the three-child families, the number of different combinations becomes 24 (3x2³), eight for each birth order (Table A2).

[3] Z=instrumental variable: same sex of the two oldest children; G=girl; B=boy

Table A2: Possible siblings sex compositions and instrumental variable value among three-child families[4]

Case	Firstborns (1^{st})	Z	Second-borns (2^{nd})	Z	Third-borns (3^{rd})	Z
1	GGG	1	GGG	1	GGG	1
2	BBB	1	BBB	1	BBB	1
3	GBB	0	BGB	0	BBG	1
4	BGG	0	GBG	0	GGB	1
5	GBG	0	BGG	0	BGG	0
6	GGB	1	GGB	1	GBG	0
7	BGB	0	GBB	0	GBB	0
8	BBG	1	BBG	1	BGB	0

In this case, some combinations cannot be distinguished. When we look, for example, at Cases 5_1^{st} and 6_1^{st}, we find the situation is as follows. The reference individual is a first-born girl who has a brother and a sister. Since the information on the birth order of the next children is missing, Case 5_1^{st}—in which the girl has a brother immediately after her, and then a sister (i.e., the two oldest children are of mixed sex)—cannot be distinguished from Case 6_1^{st}– in which the girl has first a sister, and then a brother (i.e., the two oldest children are of the same sex). The same problem arises when the observed individual is a first-born boy with one brother and one sister (7_1^{st} and 8_1^{st}), as well as when he/she is the second-born with a brother and a sister (Cases 5/6/7/8_2^{nd}). The situation is different when the third-born is observed: even if the four cases are not completely identifiable, they all take the value zero with regards to the instrumental variable. Nonetheless, French data enable us to identify one of the three groups that are counted together, namely, the second-born group. Indeed, the French questionnaire does not ask the respondent directly for his/her birth order, but rather for the number of brothers or sisters born before him/her. Thus, if the individual is the second-born and he/she has one brother and one sister, it is possible to make a distinction between cases in which the instrument is zero or one (i.e., Cases 5, 6, 7, and 8 are all identifiable for the second-born).

[4] Z=instrumental variable: same sex of the two oldest children; G=girl; B=boy

However, this does not help us in distinguishing the "mixed" category among firstborns (Cases 5/6/7/8_1st), or in solving the problem for the Italian data. Thus, the solution proposed here assumes the following.

As far as the Italian data are concerned, the distribution of each birth order among families with three children of the same sex can be assumed to be equal to the distribution of each birth order among families in which the first two children are of the same sex, and the third is different. In other words, the hypothesis is that the total number of GGGs (regardless of the birth order) equals the sum of GGBs (Cases 1_1st + 1_2nd + 1_3rd = **6_1st** + **6_2nd** + 4_3rd). Since the three GGG cases are countable, it is easy to obtain each GGG's birth order proportion on the sum of the GGGs. Further, knowing the GGB_3rd number (Case 4_3rd), only two cases remain undefined. Thus, in order to establish the proportion of the unknown GGB_1st and GGB_2nd, the assumption that assigns the (known) proportion of GGG_1st to that (unknown) of GGB_1st (and that of GGG_2nd to that of GGB_2nd) does not appear to pose too many risks. The same holds true when boys are examined.

For the French dataset, the solution is even easier to find. Indeed, by maintaining the distinction between the frequency of same sex three-child families and the number of families in which the first two children are the same sex, the unknown cases collapse into one. In other words, assuming that 1_1st + 1_2nd + 1_3rd = **6_1st** + 6_2nd + 4_3rd, only Case 6_1st has to be estimated, and it can be simply derived from the algebraic sum of the known terms. The same procedure can also be applied to the boys' cases.

Henceforth, the combinations for which the two oldest siblings of the same sex instrument takes the value one can be randomly imputed in each sibling's group, following the distribution obtained in the previous stages.