SciencesPo

Séminaire

Les sciences sociales en question :
grandes controverses épistémologiques et méthodologiques

Compte rendu de la 71e séance

When social sciences meet genomics

22 April 2024

Nonna Mayer

Today our methodological seminar is about the relationship between social sciences and genetics, and beyond about nature and culture, genes and the environment, an old debate.

For years, social sciences were extremely reluctant towards genetics, especially at the time when there was the so-called “gene of participation” or “gene of Democrat and Republican support” in the US, but in the last years, there has been fantastic progress in genomics, in the sequencing of the genome, and more and more people are using it, even on the side of sociologists. So to look into closer, look into the relationship between social sciences and genetics, we have today two specialists.

On my right, Franck Ramus, a cognitivist, who works at the Laboratoire des Sciences Cognitives et Psycholinguistiques (ENS/CNNRS), and studies children cognitive development, among other things[1].And on my left, a sociologist, Zachary Van Winkle (Sciences Po/CRIS), interested in life courses and impact of family transitions. He also has tried genetics[2].
The floor is yours, the idea is that each of you tells us from his own disciplinary point of view what one can do with genetics and then we'll have a general discussion.

Franck Ramus

Within social sciences, I am going to talk about what's closest to my own interest, which is educational outcomes. But I take it as an illustration of the contribution that genetics can make to social science, and I'm sure that Zachary will show other social outcomes of interest. Of course, I am perfectly aware that some people have very strong objections to this kind of research, and to any discussion of the role of genetics, but in the interest of time I have chosen not to address them at the start. Of course, feel free to raise them if necessary, I will be happy to discuss them.

Let us begin with what I call the greatest result of the sociology of education, which is the influence of the family's socioeconomic status on child educational achievement. I'm showing here one particular result, which comes from the PISA studies conducted by the OECD, but basically any study that collects information about both the family background and educational outcomes finds this relation. The correlation is about 0.3-0.4, and it is probably the most replicated result in the sociology of education.

This result is widely interpreted as reflecting a causal effect of the environment

[1] See Franck Ramus, “The influence of sibship composition on language development at 2 years of age in the ELFE birth cohort study, *Developmental science*, 26(4), 2023 (with Gurgand, L. et al.; “Brain volumes, thicknesses, and surface areas as mediators of genetic factors and childhood adversity on intelligence, *Cerebral Cortex*, 33(10), 2023, p. 5885-5895 (with Camille Williams and Hugo Peyre); « Ethique et génétique », *Ramus Méninges*, https://ramus-meninges.fr/2020/12/11/ethique-et-genetique-2/

[2] See Zachary Van Winkle, "Genome-Wide Heritability Estimates for Family Life Course Complexity", Demography, 58 (4), p. 1575-1602 (with Dalton Conley); "Does Parental Separation Lower Genetic Influences on Children's School Performance?", Journal of Marriage and Family, 83 (3), p. 898-917 (with Tina Baier).

provided by parents on the educational outcomes of children. Of course, it's a very plausible interpretation, and nobody really disputes it. Yet, I would like to have you notice that it's only a correlation, and if you're a social scientist who worries about distinguishing causation and correlation, well, you may want to be a little cautious. In particular, if you are attentive to the difference between causation and correlation, you should worry about confounding factors. I am sure that in your own studies you do.

Here, basically, we have a correlation between some characteristics of parents and some characteristics of their children. The elephant in the room is that there is a major confounding factor: the genome, because parents have a genome that influences their own outcomes, and half of it is transmitted to their children where it may produce similar effects, thus yielding a similarity between parents' and children's outcomes, independent from the environment. This implies a second potential causal path from parental characteristics to children's characteristics, which means that this correlation may indeed reflect some environmental causation, but there is a major confounding factor, which is genetic. If we don't take the alternative genetic causal path into account, we don't really know if an environmental causation is proven.

So that's the first caveat I want to raise, which is that in social science, we sometimes jump from correlation to claims of causation without controlling for genetic factors. Such a jump is not always legitimate.

In terms of educational outcomes, I'm going to talk mainly about educational attainment: this is the final outcome of the educational process, like the highest degree that people earn. That's a very simple measure to make. But of course, you can also replace this by any other measure of educational attainment or performance. Any score in an exam or in a test would do, the same results generally hold across all educational measures.

A child's educational attainment is obviously influenced by its own individual traits such as cognitive abilities, as well as motivation, behavior, effort, and so on. If you look one step upstream, what are the causes of this child' traits? There are two classes of causes: genes and environment.

So in principle, if you have sampled all the possible ultimate causes in the child's genome and in the child's environment, you have all the potential causes of their cognitive and behavioral traits, and ultimately of educational attainment.

Now if you look at the child's environment, you know that it is shaped to a large extent by parents, and it is influenced in particular by parental education and by parental resources. We usually collapse those two measures (sometimes together with professional category) into a variable that we call socioeconomic status. Socioeconomic status is well understood to shape the child's environment and to have an influence on their educational attainment. As I said, the size of the correlation is between 0.3 and 0.5.

On step further upstream, parents' socioeconomic status is influenced by their own cognitive abilities, and those abilities are themselves influenced by their environment and by their genes, which are transmitted to their children. That's the second causal path. Thus, whenever you measure this correlation between SES and educational attainment, it could be mediated through the environmental path, but it could also reflect the confounding through the backdoor path, through the parent's genes that influence both the parent's socioeconomic status and the child's genes and therefore their cognitive abilities and educational attainment.

It turns out that some genomic studies have suggested that this backdoor path accounts for basically 50% of the SES-educational attainment correlation. This does not challenge the idea of causation through the family environment, it definitely exists, but it may be overestimated by the raw correlation. If you properly adjust on genetic transmission, then you get a revised, lower estimate of the environmental path. Given that the genetic mediation accounts for about half of the effect, it's pretty important to take it into account.

I have given away one of the main results, but I'm perfectly aware that you don't have to take my word for it. You are entitled to ask for some data before you believe me, and the time is very short so I'm not going to be able to show you all the relevant data, but I'm going to refer to the main sources of data and to show you some pieces, with pointers for those who want to go further.

Why should you believe that there are genetic influences on educational attainments? Broadly speaking, there are three main sources of evidence.

The first type of evidence is family studies, the most well-known ones are the twin studies that have been going on for almost one century, but also adoption studies, and there are also other kinds of designs using family data. The second kind of studies is genetic mutation studies, and the third kind is genomic studies. I'm going to illustrate each of them.

First, family studies. As you know, there are identical (monozygotic) twins, which are genetically identical, and fraternal (dizygotic) twins, which are only 50% genetically similar. Those bars show the correlations of scores in various tests or questionnaires within twin pairs: monozygotic twins in black and dizygotic pairs in hatched bars.
As you can see, the black bars are always taller than the hatched bars, so the identical twins are always more similar than the dizygotic twins. This is generally interpreted as showing evidence for genetic influences and this is used to compute what we call heritability.
Note that some of these traits are relevant for educational outcomes, such as general intelligence, verbal reasoning, memory. There is also scholastic achievement in adolescence, so this must be a grade in an exam in high school. There is even vocational interests.
This graph shows you the same result as in all twin studies, that anything that you can measure in an individual seems to be heritable to some extent, including cognitive and educational outcomes.
I'm aware that twin studies have been heavily criticized, we can go back to that if you want, but it's important to know that they are not alone. First, they have been very well replicated. Secondly, they are complemented by adoption studies which have different assumptions and methods but that provide very similar results.

Now, all those family studies allow us to quantify heritability, which means that genetic variations are involved, but this is very vague. Once you know that there are some genes involved, you would like to know: which ones and how do they operate?

Our initial understanding of genes was largely derived from studies on genetic mutations. This can be traced back to the identification of the genetic basis of Down syndrome, which is caused by an extra copy of chromosome 21.

While this is a clear example of a genetic mutation, there are many other less dramatic mutations that have been discovered since then. Currently, we know of over 1,000 genes, out of a total of approximately 25,000, whose mutations can lead to some form of intellectual disability.
It's important to note that these mutations are not limited to causing intellectual disabilities. They can also lead to other disorders, such as language and reading disorders. However, these mutations are quite rare, affecting roughly 1% of the population.

Interestingly, many of these genes have been discovered in what we refer to as multiplex families. In these families, multiple members across generations are affected by a particular disorder. Men are represented by squares, women by circles, and those with the disorder are marked in black.

For instance, consider this family where the grandmother had a language disorder. Four out of her five children, and about 50% of her grandchildren, also have this language disorder. Similarly, in a French family I studied, the grandfather had dyslexia, a reading disability, and eight of his 11 children and about 50% of his grandchildren also have dyslexia.

When studying such multiplex families through genomics, we may, with some luck, find a mutation that can be reliably associated with the transmission of the disorder. This method has been used to discover numerous genes associated with cognitive traits. However, it's worth noting that this only applies to a small fraction of people who have language disorders or dyslexia. Most of them do not carry any identifiable mutation (yet).

Finally, it goes without saying that individuals with intellectual disabilities, language disorders, or reading disorders often face challenges in educational settings. In essence, all these genes contribute to variations in educational attainment.

While rare mutations have been the focus of much genetic research, it's important to consider the broader population. Specifically, we must ask whether more common variations in the genome could explain the more common variations in educational outcomes among people who don't have an obvious disorder and deleterious mutations.

To answer this question, we turn to whole genome analysis. I'll illustrate this with a Genome-Wide Association Study (GWAS) from a 2022 paper. In a GWAS, DNA samples are collected from participants, typically via saliva or blood samples. In the study I'm referencing, they had an impressive 3,000,000 participants. This was achieved by merging previous genetic studies, regardless of their outcomes. Most genetic studies are conducted to investigate the genetics of cancer, diabetes, psychiatric disorders, etc., and most participants are also asked about their highest degree. By reusing and pooling this data from all possible data sources worldwide, they managed to conduct a study on the genetics of educational attainment with 3,000,000 participants.

The genome is screened at about 500,000 to 1,000,000 sites. These are Single Nucleotide Polymorphisms (SNPs), sites in the DNA where you can have any of the four nucleotides: ACGT. The association of each of these SNPs with the outcome is tested statistically. For each of these SNP, we know whether it is statistically associated with the outcome, and we know the effect size.

Given the difficulty of handling 1,000,000 variables, researchers calculate polygenic scores (PGS, GPS, or PRS). A polygenic score summarizes all the information across all these DNA polymorphisms. For each site, the effect size of this SNP on the outcome is calculated. For example, here, having a C instead of an A modifies educational attainment by -0.02 standard deviation, and so on for each site.

Every person's GWAS results are taken, and if they have one copy of C here, their score is incremented by -0.02. If they have two copies of G there and the effect size of G is +0.01, then their score is incremented by 0.02, and so on. In the end, you sum the effects of all the SNPs and you obtain the polygenic score, which quantifies the overall genetic predisposition of the individual for this outcome.

In general, the polygenic score predicts much less variance than all the SNPs in the genome, but still, you can predict a significant amount of variance, like 5% of the variance, or 10% of the variance, or even more if you're lucky. In this study, they obtained a polygenic score that explained between 12 and 16% of the variance in educational attainment, which is considerable.

To illustrate what this kind of effect size translates into, consider the prevalence of grade repetition. Pupils who have the lowest 10% polygenic scores have about a 30% risk of repeating a grade, whereas those in the top 10% of polygenic scores have about a 5% risk of repeating a grade.

Knowing your polygenic score does not really determine your educational attainment—it's not a very good individual prediction. But if you average over deciles of the population, you can see how substantial the effect is. Similarly, if you look at the prevalence of high school completion, it's about 70% in the first decile, and more than 90% in the 10th decile. If you look at college completion, it's about 5% in the first decile, and 60% in the 10th decile. So, polygenic scores do make a difference.

Some people might argue that a polygenic score explaining only 12% of the variance is negligible. However, when compared with the variance explained by well-known social variables, such as mother's and father's education, which each explain 15% of the variance, or household income, which explains about 7% of the variance, it's clear that your polygenic score, whether it explains 10%, 12%, or 16%, is in the same ballpark. Therefore, it should be taken as seriously as any social variable considered important.

Here is a study that examines the combined impact of father's income and the polygenic score for educational attainment on college completion. The data is represented on a graph with two scales on the x-axis: the four quartiles of father's income and the four quartiles of the polygenic score. As expected, a higher father's income correlates with a higher rate of college completion. However, within each income quartile, the effect of different polygenic scores is also evident.

This demonstrates the additive effects of father's income and the polygenic score. It underscores that there is no conflict between genetic and social effects. If you measure and analyze both of them, you can see that both are present and jointly contribute to the outcome.

Now this new study explores potential applications of these findings. One such application is the current hot topic of the impact of screen exposure on child cognitive development. The prevailing hypothesis is that screen exposure negatively affects cognitive development and educational attainment. Numerous studies support this hypothesis, showing a negative correlation between screen exposure and these variables.

However, alternative hypotheses exist. One is that children with certain cognitive abilities may be more or less attracted to digital devices, implying a reverse causation. Another hypothesis posits the existence of confounding factors, such as socioeconomic status. Differences in screen exposure could be attributed to differences in family environments and regulations around screen time.

If these confounding factors are not accounted for, the correlation between screen exposure and cognitive ability could be misinterpreted. Most sociologists are aware that, in order to correctly interpret the data, it is important to control for socioeconomic status and related factors. But these are not the only possible confounds. Children who are highly exposed to screens may also be genetically different from those who are less exposed. Therefore, it would be prudent to control for genetic transmission.

Unfortunately, most studies do not. However, one study by Swedish researchers does. They used the ABCD cohort of American teenagers, and measured three types of screen exposure: watching videos, socializing on social networks, and gaming. They also collected cognitive and educational outcomes and genetic data.

Their findings replicated the usual negative correlation between screen exposure and cognitive ability. However, they also found that socioeconomic status and the polygenic score for cognitive ability were negatively correlated with screen exposure. This suggests that both variables should be adjusted in any causal model.

In their study, they built a causal model to predict the effects of different measures of screen time on intelligence changes between two time points. After adjusting for both socioeconomic status and the cognitive polygenic score, they found that the originally negative correlation actually became positive!

This is just one study and it needs to be replicated. However, it serves as a reminder of the importance of adjusting on confounding variables, including genetic ones. Failure to do so could lead to incorrect conclusions. This is just one example of how these findings can be applied, and there are certainly other potential applications to consider.

Let me give another hypothetical example, where you would like to compare the effects of schooling in private or alternative schools like Montessori, and in standard public schools. Of course, school types may significantly differ in various measures of pupil success, but you are well aware that there may be confounding factors. It would not be appropriate to estimate the effects of private schooling without first controlling for socioeconomic status.

Again, in such a situation, genetic confounding should also be considered. The children attending public and private schools may have different genetic predispositions, which could influence their academic performance.

Thus, even in interventional research, such as randomized control trials (RCTs), it may be important to ensure that the two groups being compared do not differ significantly in socioeconomic status or genetic predispositions. These factors could impact the results of the RCT.

There may also be interactions in the statistical sense. For instance, while socioeconomic status and polygenic scores may have additive effects, this may not always be the case. Many educational and social interventions seem to work on average, but they have variable outcomes. Not all students respond to the same pedagogical interventions, and not all dyslexic pupils respond well to speech therapy.

One hypothesis is that genetic predispositions may explain why individuals react differently to certain environmental factors or interventions. Therefore, it may be beneficial to use genetic indices to understand why people react differently to these interventions.

In conclusion, there is no opposition between social and genetic factors. The genome is a major confounding factor of social influences, and it is important to consider this in any study observing correlations between parents and children, as well as in group comparison studies. It is possible to control for genetic transmission in studies of social factors with the right kind of data and design. Furthermore, the genome may also interact with social factors, which expands the range and complexity of the causal models that we should consider.

With the decreasing cost of genotyping and the increasing availability of databases containing data on children or adults, including genetic data with social and educational outcomes, it is becoming more and more feasible to answer questions with proper genetic controls. Even if you don't collect such data yourself, you may be able to find the right database to answer your question.

Zachary van Winkle

Thanks for the invitation and thanks for the great introduction, because I took a bit different approach when thinking about what to talk about. I decided to talk as a specialist of the sociology and the demography of family, and tell you a little bit about how I and my colleagues apply socio-genomic approaches in the field. A lot of the methodological approaches have already been introduced here, which is quite nice. What I did is basically pick out two areas of my research where I've taken a socio-genomic approach and these two areas also have questions behind them that might be interesting for social scientists here.

The first one concerns socio-genomics, do we actually need it, or can we ignore it? The first time that I started reading about SG as a family demographer, I would say around 2017, I thought, ideally, we would find out we can ignore this. As you can figure

out with the next question, it turns out that we probably can't ignore it. So how should we use it in our research? and I thought I'd try and give an example there.

This first project was published in *Demography* a while back, and it's entitled “Genome-Wide Heritability Estimates for Family Life Course Complexity”[3]. I assume that most of you have not gone through all of my publications and might not be familiar with the exciting world that is family life course sociology, but one thing that I and my colleagues are interested in is this sort of dramatic increase in the complexity of family life courses and family formation across the 20th century. When I talk about family complexity, I mean is increasing number of transitions in the life course, people don't just get married today and are done. They cohabit, they get married, some don’t, they get divorced, maybe they get remarried. Things are a bit more complex in that sense, also a bit more unpredictable.

There are a lot of sociological and demographic theories to explain why that increase has taken place and why there are cross-national or cross-regional variations in complexity. The most prominent one is probably the Second Demographic Transition thesis. Across the 20th century something arose, that we might call post-materialism, people became a bit more liberal in their values and decided, for instance, it's acceptable to have children outside of marriage, I want to have families in my own way as a route for self-fulfillment and so on. But what wasn't really considered in life course sociology at the time, and this has changed slowly, is what I'm calling biodemography, but what we might also call socio-genomics. The idea of incorporating biological approaches with our sociological, economic, demographic theories. And these approaches can be both methodological and theoretical.

The first step to taking a biodemographic or a socio-genomic approach, I would argue, is what we already heard about from Franck’s presentation, it would be to simply estimate the heritability of the outcome that we're really interested in. Heritability in this sense is going to be something like the proportion of the variance of a trait that's attributable to a genetic variance. If something is not heritable at all, then we might consider ourselves lucky as social scientists and just ignore this, but as we saw earlier

[3] Zachary Van Winkle & Dalton Conley (2021). *Demography* 58 (4): 1575–1602.

with this nice graph on twin studies, it’s probably unlikely.
But considering family life courses complexity, we thought we would give this a try and look to what extent this complexity is heritable. We also wanted to test a key theory in family demography and see whether we observed cross-cohort variation in that heritability.

Why should outcomes be heritable, especially family demographic outcomes? There are probably many direct mechanisms that link genomes or genetics to our outcomes.

So if we think about menarche -the first occurrence of menstruation - or simply fecundity as the biological capability to have children, these are probably directly impacted by genes. But there are also numerous, plausible ways where there are indirect pathways between the genome and the trait that we're interested in what has been called “intermediate phenotypes”.

Previous research in family demography that used twin studies showed that a lot of the components of family life course complexity are heritable, with high estimates for instance for women's age at first birth, total number of children, propensity to marry, and to divorce.

Now like we heard, there's been a recent shift since the sequencing of the genome towards heritability estimates that are based not on assumed related genetic similarity (twin studies or family studies), but on measured genetic similarity, usually through genome- family demography for genome-based heritability on family demographic outcomes, really only for completed fertility and the age at first birth, both of which were considerably lower than the twin estimates.

So on the one hand, with this study, we wanted to update these findings, have more genome-based heritability estimates, and on the other we wanted to test a key hypothesis in family demography, basically proposed already in the mid-90s, that the heritability of family demographic outcomes should increase over time as societies open and become more liberal, and behavior becomes less constrained by societies.
For instance when divorce becomes easily accessible for everyone, then obviously any sort of genetic propensity has the possibility to emerge. And at the time there had really

been very few empirical tests of this hypothesis, mainly only with twin studies.
So we used a dataset that has already been mentioned briefly, the United States Health and Retirement Study. It is a Biennial Panel Study of Older Adults, that began in 1992. As individuals are followed up every two years, there's a refreshment sample that's collected about every six waves, and we have pretty good information on family demographics, so we have retrospective and prospective information on marital histories and fertilities.

This allows us to reconstruct the entire family history of an individual, whether they are single, married, divorced, how many kids they have, at what age etc. And to calculate a commonly used indicator in family life course sociology called the complexity index, our most distal outcome if you want to capture core family life course complexity. On the other hand we have is genetic information, for about 15,500 respondents, we have about 2.4 million SNPs (Single nucleotide polymorphisms, the most common type of genetic variation among people).

Then we calculate a genetic relationship matrix, for all non-Hispanic whites in our sample, and we basically , based on those SNPs, how genetically similar everyone is in our dataset, so this is basically a pairwise similarity matrix, and we perform something called a “genome-related evidence-based restricted maximum likelihood model”(GRML), to estimate genome-based heritability for our outcomes.

Some of the social scientists here may have used random effects regressions to estimate inter-class correlation coefficients, where we say the proportion of variance is so-and-so much attributable to country differences or school differences? This isn't that much different, but here we're not looking at country differences or schools, we're looking at genes.

What we have in the main results are a wide range of strange indicators.

First complexity indicators: the fertility complexity index, and the differentiated sequence complexity index, commonly used in family life sociology. Here you see something that's a bit strange with the HRS data: in this sample we don't observe an increase in family complexity, , because our youngest cohorts are just at the start of

the second demographic transition, so this sort of emerging family behavior is just being captured, as you can see by the increase in variance by our youngest cohorts. Nonetheless we went ahead and looked at heritability. And so for example on the top left there for fertility complexity, we see on average that somewhat less than 20% is attributable to common genetic variance, and that seems to be a bit higher for our older cohort compared to our lower cohort. We see that trend is similar for other indicators of distal life course outcomes, such as the age of first marriage, or the propensity to divorce, that there is non-zero heritability. This is something that should interest life course sociologists, but also family demographers and family sociologists.

One thing that we didn't find though is evidence for increasing heritability, like I said this could be due to the HRS sample, bit of a strange, maybe a bit too old of a sample for this, but this is something to keep an eye out for because we have so little data to be able to do this, we must wait until we have DNA samples for younger cohorts.

Now to move back a little bit to education, one area of research that I've been working on is looking at parental separation and how that impacts genetic influences on education, getting towards the idea that okay, socio-genomics is important, now what do we do with it as sociologists? We know there's a pretty established link between family of origin and children's education. We also just heard that not only social but also genetic factors are important for education and its predictors, and also that genetic influences can vary according to social conditions. So a little bit of thinking is needed about gene-environment interactions.

Most of the research on gene-environment interactions often focuses pretty narrowly on IQ, sometimes other dimensions such as educational attainment, but also pretty narrowly on parent socio-economic status as measure of the children's environments. This is often referred to as the Scarborough hypothesis, the idea that the relative importance of genes for IQ or for other school outcomes should increase with parents' socio-economic status, and the relative importance of the environment should decrease. Because in high SES households, parents are investing in their children quite specifically, these are enhancing environments.

Talking to my colleague Tina Baier at one point, I was thinking that parental SES is a bit limited because one of the big emerging dimensions of family advantage and disadvantage, at least in the United States, is divorce and separation, and this isn't necessarily captured by socio-economic status alone.

We know, of course, that there is socio-economic selection into divorce and parental separation, and that they also induce socio-economic hardship. But children, regardless of their socio-economic status, are exposed to increased stress during separation processes, conflict, parental absence, as one parent simply has less time than two parents. So parental separation also creates distinct family environments that may lower children's chances for genetic, to realize their genetic density for education. Most of the literature in this area is looking at the direct effect of parental divorce on educational attainment, we're swapping that a little bit, looking at sort of the moderating effect, whether parental divorce suppresses the genetic potential of children's genes for educational attainment.

In the first initial study using the German twin life data, we looked at a number of outcomes, such as the results for cognitive skills for children, and we used a decomposition method where cognitive skills for one parent and two parent households are decomposed into proportions attributable to genes, the shared environment, like the family, and then the unique environment.
This is done contrasting mono and dizygotic twins, and we see, as expected, that the genetic component here is larger for two parent families compared to one parent family. In this study we could additionally adjust for mother's education and household income, it does not change the results. But we couldn't go much further, we couldn't, for example, look at impact of stress, conflict, parental absence.

What we're currently doing now, is switching the context, getting away from Germany and going back to the United States. In this study we're bringing in a third reference group, a group of individuals who experienced parental death during childhood, and not just stable, two parent family. Parental death in a way may be more similar to stable two parent households, because there's less selection into death compared to divorce, and there's really not much reason to expect a pre-death increase in conflict, most deaths for these cohorts are accidents during these ages, nor an increase in care work

prior to death. But of course our parental death group could be more similar to our divorce group if it's not conflict, if it's not selection that's creating change, if it's really the absence of a parent.

So we use the same data that I just talked about, the US Health and Retirement Study (HRS), but we also link information from the 2015 and 2017 Life History Surveys, where we have a lot of data about childhood living circumstances, including whether individuals grew up with both parents or not. The main dependent variable is simply the years of education.

We include one of these PGSs that was already explained, the Education Polygenic Score available in the scientific use file, so everyone here can access it, thus we have a childhood family structure variable with parental divorce, death, or two-parent family. And we calculate OLS regressions where we interact those two, to see whether the PGS has a different impact on education depending on family structure. We also adjust for something called the First Ten Ancestry Principle Components (we can talk about it later) , as well as gender and year of birth.
And, also something to talk about maybe later, we exclude individuals of African ancestry, only including European ancestry respondents in the HLS here.

The main finding is, again, that genes are more important for educational attainment in two-parent families, compared to separation. One standard deviation increase in the PGS is associated with about a 0.61 increase in years of education, compared to about a 0.38 increase for those who experienced a parental divorce or separation before the age of 16.

Then we wanted to know if it was the effect of selection, or economic hardship? We have some indicators such as parental education, self-rated socioeconomic status, etc,. We included them as well as interactions in our models to see whether we could attenuate that interaction effect. We couldn't. So it doesn't seem to be socioeconomic selection or hardship.

Now, when we do this indirect test for the next mechanism, stress, we don't see statistically significant differences in the interaction between the parental death group

and our family structure, our two-parent family group, but we do have a statistically significant, large, negative interaction for our divorce group. This suggests that our finding is probably driven mostly by family conflict, by stress, although it's important to remember this is an indirect test.

The next step is something we always want as family sociologists, we want more information, direct information on family circumstances.

We technically have this in Add Health (National Longitudinal Study of Adolescent to Adult Health), we just haven't gotten access to the data yet.

Christiane Silliau

I'm a biologist. I just wanted to make clear that now genes and genome are considered in a very different way because we know that genes are only, for human beings, 2% of the genome. So the first experiences you show, they were with genes, but there was something when you went from one nucleotide to another independently and you calculate the percentage of each one and then you summarize this, I want to know how this test has been done, considering I suppose the nucleotide in general, that means the whole genome, rather than genes, no? Because you count even millions of them, so I suppose...

Franck Ramus

Yes, so those genome-wide association studies are based on the SNPs that are carried by commercial DNA chips, and those DNA chips focus on the single nucleotide polymorphisms that are the most variable in the human population. I don’t know to what extent there is a selection for those SNPs that are included in genes, but the SNPs included are definitely not restricted to genes. The fact that SNPs may be outside of genes is not necessarily a problem, because these may be regulatory regions, where other molecules bind to the DNA in order to regulate nearby genes.
Zachary do you have something to add?

Zachary Van Winkle

No, I think that's a good point. I mean, I'm a sociologist, and so my understanding is also that like the HRS and other data sources that we have with these SNPs, the selection made is based on the idea that these are the most common areas where there's variation in the population, or where there's a specific interest. And I guess what you speak of, that's one reason why we would expect twin studies, for example, to have higher heritability estimates than estimates that are based on Grimmel or PGS. So it's basically the methods that are based on these 2.7 or 2.5 million, however many SNPs.

Jens Carstens (CEE)

I've got two very big questions. First, what is the role of race and why do you exclude people of African American descent? And the second is about data access, as I'm working on political behavior and political attitudes. Are you aware of any data sets that include these type of variables?

Franck Ramus

Regarding race, here is the main problem: there are many genetic variants that occur more frequently in certain populations than others. For example, a number of SNPs differ in frequency between European and East Asian populations. There are also cultural traits that differ, for instance eating with fork and knife vs. eating with chopsticks. Thus, these genetic differences are statistically associated with cultural traits, yet you wouldn’t want to conclude that you have identified a genetic cause of proficiency with chopsticks. You know that these statistical associations exist because, in the course of human history, these populations have simultaneously genetically diverged and culturally diverged. So, more generally, whenever you find statistical associations between genetic variants and traits *across* populations, you don’t know how to interpret them: they may reflect causal effects but they may also be historical coincidences.

That’s why genetic studies tend to restrict their analyses to relatively homogenous populations, which may lead them to exclude minority populations from certain

analyses. Furthermore, they use various statistical techniques to try and account for population stratification, even within a given geographical or ancestry group.

Regarding genetic studies of political attitudes, this is really not my area. But I am pretty sure that there are databases including both genetic information and questions about political attitudes, and that such studies exist. Zachary may know more.

Lydia Panico (CRIS /Sciences Po)

Thank you very much for these presentations. They were really fascinating. I have a kind of follow-up question. You talked already of the fact that often people of African ancestry are excluded. I also noticed, like you mentioned, the ALSPAC (Avon Longitudinal Study of Parents and Children) which we know is quite an advantaged cohort. They are from a part of a suburb of Bristol, if I'm not wrong, that's quite an advantaged suburb. And I don't know whether it is actually used, but for example, I know the genetic biobank in the UK is also a very advantaged data source because you have people who are willing to give up the time to go there every, I think it's a year or so, to have blood taken. I think they do MRI scans and so on.
You also mentioned that perhaps as people are more advantaged, gene expression is easier.
And so I'm just wondering to what extent the fact that you are estimating these analyses on quite advantaged samples overestimates the role of genetics in this?I don't know if my question is clear.

Franck Ramus

I think it's entirely true that in almost all data sets in the world, research is based on volunteers, therefore the population is biased towards higher socioeconomic status. Still, all parts of the population are represented, but some may be less represented than others. I think the main problem is that population variance is restricted, and that this reduces statistical power. I'm not aware that this induces major systematic biases. It may induce a bias if indeed there were different genetic effects at the bottom of the SES scale as opposed to the top.
Do you want to add something?

Zachary Van Winkle

I think it's something interesting to think about.
But it wouldn't always necessarily be that heritability is being sort of overestimated. It could also be underestimated depending on whether you expect that there's more of an enhancement or a suppression versus a triggering effect. You might think that maybe genetic effects are larger among disadvantaged groups, for example.

Tina, my co-author, is also working on mental health outcomes. And there with parental separation, this is thought to be a triggering instance that may induce the expression of certain genetic predispositions.

Franck Ramus

I suppose an even more general answer is that the problem might arise if there are genuine gene-environment interactions (non-additive effects). Interestingly, twin studies have found many gene environment interactions, such as the Scarr-Rowe hypothesis shown by Zachary. So far, they have mostly failed to replicate in genomic studies. Thus, when people have tested the potential interaction between a polygenic score and an environmental factor, they have generally failed, for reasons that we still don't really understand.
So, to the extent that we don't find such interactions, it doesn't really matter whether you sample isat the top or the bottom of the SES scale. But of course, we should keep this in mind as a potential bias to be assessed.

Zachary Van Winkle

Yes, maybe something that's interesting with this is that the first study that Tina and I did with twin studies was published quite quickly and easily. We were actually surprised how gentle our reviewers were. This journal of family sociology, the Journal of Marriage and Families, is a quite well-known journal. Yet, we had quite an easy time publishing it.

But the current version of the paper that I presented with the polygenic risk score, I think we've now been rejected three times and we generally have a reviewer that is quite clearly more of a ” socio-genomicist”, someone a bit more on the genetic side of things.

My impression is that there's more an idea that gene-environmental interactions, which we as sociologists may find substantively interesting, are often seen more as a methodological issue in the field. And so this is something where, when it's possible, it's nice to have chats with them to hash this out a little bit. But I don't really have too many answers for you there.

Claire Andrieu (Sciences Po, Centre d’histoire)

I have two questions.The first one is for Frank, who evokes strong opposition to mixing genomics and sociology or social sciences. I would like to know more about those strong oppositions.And another question, did you work on counter-examples?For instance, people with identical environment and identical PGS who have very different educational achievements, or vice versa?

Franck Ramus

I’ll start with the last question, which is easier.
What comes immediately to my mind is that people have studied monozygotic twins, who are genetically identical, and who are raised within the same family. They are extremely similar, nevertheless, they are also to some extent different, and they are sometimes discordant: one has a disorder and the other has not, or they don't have the same level of achievement.
Such case studies may help identify some of the factors that may cause a difference between twins. In pregnancy, one may grow larger at the expense of the other, so there can be asymmetries. Then, in the course of life, many different things happen. For instance, one can have an accident and the other not. So even strictly identical people can become different to some extent. And of course, they may not be absolutely genetically identical: one may carry a mutation that the other one doesn’t have.

Considering opposition to genetic research, most of the opposition I hear is people who are worried about the potential implications of genetic results. These debates have been ongoing since the 70s, beginning with the sociobiology debate.
There are different kinds of implications that people worry about.
Many people have reservations about genetics because they worry it presents an overly reductionist view of humans. However, as we've shown, the idea is not describe humans at a purely molecular level. Understanding of genetic causes can and should be integrated with other levels of explanation (cellular, physiological, cognitive, phenomenological, sociological…).

People often perceive genetics as deterministic, but in reality, every factor that scientists study is deterministic to some degree. Genes are not more deterministic than any other factor.

Yes, people are apprehensive about the potential applications. How are some going to use this genetic data? It could be used to discriminate or to select. People are indeed very concerned about these potential applications. And they should be. As they should also be concerned about the applications of all scientific knowledge. Medical knowledge, for instance, can be used to save lives or, conversely, to end them.

So yes, we should always be concerned about applications of scientific knowledge. We already have bioethics laws that regulate what can be done with genetic information. However, we should continue to be vigilant.

Dominique Turcq (Boostzone Institute)

I'm a prospectivist working mostly on the implications of what you are debating here. The question is about China, you mentioned mostly research from the US and from Europe.
Obviously, China is researching a lot on this subject. Have you any information of where they are compared with Europe? Do you work with them? Do you know them? And the second question is about the implications. Do you have already contacts, both of you, with people from the management area, from the legal area and so on, asking you about some of the implications you may see for them in these areas?

Franck Ramus

About China, until recently they were lagging behind in terms of genomic studies. But they are, as you can imagine, catching up quite quickly. For instance, I'm part of a language genetics consortium, of which they were initially absent, but now they are contributing more and more data.
It's important for every region in the world to contribute, because until now, most studies have been based on European populations, which gives a biased view of genetic diversity (hence the race question previously addressed). We have seen indeed that when a polygenic score is calculated on a European population, and when you try to apply it on an African or Asian population, the prediction is much less good. Because, of course, the score has been optimized for the European population.
So it's important to make sure that every population in the world is included in such studies, in order to have more reliable estimates and to have a broader view of human diversity.
Regarding your second question, I don't think I have anything to say.

Dominique Turcq

Nobody approached you?

Franck Ramus

No

Dominique Turcq

Isn't there greater reluctance in France than in other countries?

Franck Ramus

Yes, possibly.
Thinking of what legal applications might be, you might want to use a polygenic score

to estimate, for instance, a genetic predisposition for violent crime. But I don’t think that this kind of evidence could be used in a trial, it's just not reliable enough.

Let me discuss a potential application that aligns better with my area of expertise. Some have proposed the use of polygenic scores for educational attainment in educational contexts. The idea is to identify children who have special talents or predispositions, such as for mathematics or music, or those with genetic predispositions for certain disabilities like dyslexia or ADHD. The goal is to use these predictions, which can be made from birth due to the stability of DNA, to tailor their environments and provide additional support to prevent the development of a disorder or to further develop talents.

However, I personally believe this is not really feasible. Even the polygenic scores we hope to have in the next 10 or 20 years will likely not be predictive enough. The error margins for individual prediction are so large that it's hard to imagine how one could meaningfully act on them.

Moreover, for many of these conditions, we already have predictors. For instance, for dyslexia, which I am quite familiar with, we already have predictors such as language delay, general language difficulties, and problems with verbal short-term memory. These can be detected in kindergarten, three years before the child learns to read. At that stage, we can identify children at risk and attempt to enrich their linguistic environment. Would there be any advantage in having the same information at birth, but with much less reliability? It's not clear to me
.
Similarly, if you want to detect talents in math or music, would it be more beneficial to do so at birth rather than in kindergarten? In kindergarten, if you want to determine if a child is gifted in math or music, you simply expose them to these subjects and you will quickly see their capabilities.

Therefore, it's not clear that polygenic scores are the right tools for making useful predictions in real educational settings. That's my current perspective on the matter.

Dominique Turcq

I can just add to your answer. I'm glad that you see most of the positive benefits we may have.
One of my most important concerns is about the labor market. The labor market may use genetic analysis to select people, as we did 20 years ago with graphology, which was not scientific at all. A lot of people did not get a job because they had bad writing. Tomorrow, some people may not get a job because genetic analysis may show that they are not good at learning or whatever. And here we have a major risk on the market, even if it's not scientifically valid yet.

So this is why I'm concerned about it. I can see that, and I think that it's going to be as reliable as graphology. And this is especially seen since there are already much better predictors of job performance than such tools, right? I mean, like cognitive tests, personality tests, job tests.
It will become a gadget, like the graphology.

George Marcus (Professor emeritus, Williams College)

I spent a sabbatical year at the University of Minnesota and got to know Auke Tellegen and read and discussed much of his work. I was an enthusiast then, and I'm an enthusiast now. Nonetheless, I think we're centuries away from a sound theoretically driven understanding of the roles of genomic influences on human experience. So these are things you guys know, but maybe not everyone else here.

Humans have two eyes and four chambered hearts. Variations on those are so rare, in part because they are less useful for environmental fitness and risk death..The human species evolved so that certain features are absolutely fixed. Other features are highly variable because evolution has taught our species that variation enhances species survival. Moreover, the human species spends roughly a quarter to a third of our lives in developing into adulthood. There are many mental processes that are still evolving at 17, 18, 19. This affords a considerable time for acquisition and mastery of culturally specific practices as well as maturation of neural, and other, processes.

When is understanding genomics going to prove useful in any specific social context given the enormous variation in culturally specific settings, and of the natural world and cultural settings? That's the aspiration. But until some effort is put into theoretical development all these ‘models’ are wildly underspecified. I'm not saying that as a critique so much as the steps that have to do to fill in are enormous. So for example, on family complexity. You claim, and I agree, complexity of family life and life courses has over centuries became more complex. The Enlightenment anticipates a new social order that's highly cosmopolitan, highly dynamic, encourages yet more highly complexity. And everybody who's in favor of that, of which I am one, thinks that is the new normal. But from the very beginning, there's roughly 20 to 40 percent of any population that says we hate that life. Yet, some want a life of the village. A stable life of recurring cycles. While others, most it would seem, favor the life of the modern city, dynamic, more open possibilities, and ever changing. It is likely that each person’s genetic structure place you somewhere on a dimension that ranges from favorable to the first or favorable to the second.

But as each person is born into a specific situation it maybe that your life begins in a village life and your desire is to get out as quickly as you can, if you think of the possibility of getting out, right? Or you're born better suited to be a villager and take over the family farm, but you're born to families that are living somewhere in Paris or Copenhagen or whatever, where who farms? Your parents aren't farmers. So those are the dynamics that we really, if you're a political scientist, want to get a handle on. Until you get to that level, it will be longtime before genomics can tell me about that specifically about who is attracted to the one as against the other. It’s a fine topic but offers little current political insight.

So one of the things I would suggest actually on your family study using US data, US states create enormous variation in the level of state support for families. So instead of focusing narrowly on the genomics on family structure, if you add the social dynamics of support varying at the state level from low to high, you then get a much richer set of interaction possibilities that would be really interesting.

Nonna Mayer

Do you two want to answer?

Franck Ramus

Indeed, genome-wide analyses are atheoretical, they treat each SNP as a random variation to be analysed independently of each other. In the early 2000s, there were more theory-driven “candidate gene studies”. Based on biological function and animal studies, people hypothesized that variations in one gene, say a transporter for dopamine, was involved in attention or impulsivity, and went on to test the association with ADHD and the interaction with some relevant environmental factors. It turns out that most of those candidate gene studies failed, their results subsequently didn’t replicate. That’s why people moved on to atheoretical genome-wide studies.
Nevertheless, biological understanding and theoretical interpretation can come after discovery. From the many SNPs associated with a trait in a GWAS, it is possible to study the biological effects of these variations, e.g., a change in the shape of a protein, or the modulation of the expression of a protein. And across all the genes involved, people study globally what physiological pathways seem predominantly involved, etc. This is all entirely tractable, it’s just very complicated because there are many SNPs and many genes involved, and each only has a tiny effect size.

From the studies we know, we can confidently say that there is no specific gene for people who thrive in villages as opposed to big cities. The best we can hope for is a polygenic score that would predict a decent amount of variance. And biological studies could in the long run inform us on the mechanisms that predispose to such personalities (for instance, neural factors that influence exploration-exploitation choices). However, this is going to be a long-term endeavor.

George Marcus

Let me add one more burden. With respect to personality, we name a trait with respect to one end of the distribution and then imply that one end as malignant or deficient. For example, the trait named ‘emotional stability’ was originally named ‘neuroticism’ - neither is accurate, but the normative attribution is clear as to which end of this trait is best for what is left unclear. This pattern raises a question, why does that variation then have persistent heritable effects? If heritability ensures there's variation, the presumably that variation is helpful, and that's why they're being reproduced rather

than shut down.
And in fact, the most interesting slide was the one you showed of family passage. It shows that genetic effects wash out very rapidly by the second generation. The grandparents' genetic contributions have been more than halved. Why is the genomic system seemingly designed to have short-term genomic effects that will diminish over time, presumably to make way for still yet more variations than having been born? Do you agree?

Franck Ramus

I would not say that effects wash out over successive generations. Of course, each parent's contribution is halved after each generation. But at each generation, the number of offspring multiplies. So, in fact, to the extent that your offspring are able to reproduce, your total genetic contribution to subsequent generations does not diminish, it is just spread ever more thinly across an increasing number of descendants.

With respect to traits, I think it depends. There are some traits that have a clear valence, for instance a higher IQ score is better than a lower one.

George Marcus

On what dimension of intelligence?

Franck Ramus

For all dimensions of intelligence, any cognitive ability, the distribution of scores reflects a continuum between low and high performance, and the latter is more adaptive and led to greater reproduction (in ancestral environments). So there is no question that for some traits, better performance has been selected in the course of evolution.
The fact that there is still a lot of variation in cognitive abilities does not contradict this idea. It does not imply that lower scores are also adaptive in some way, or that diversity is maintained because it would be valuable per se. Human intelligence is not fixed because it is still evolving, or was still evolving until recently and the generalization of

contraception.
Others traits are different, as you said, personality traits have no clear valence, any place in the continuum can be adaptive. In this case, diversity exists and is maintained simply because there is not one single way to be successful, even in the narrow evolutionary sense of maximising reproduction. You can maximise your descendance by being very conservative, cautious, sticking to the time-proven ways for yourself and your offspring. You can also maximise it by being very open to new ideas, to new behaviors, to new environments, and by discovering new ways to thrive. Both strategies have their pros and cons, their probabilities of success and failure, and can co-exist in a form of equilibrium.

So the fact that human traits are genetically influenced is not contradictory at all with the maintenance of cognitive and behavioral diversity, alongside genetic diversity.

Nonna Mayer

First, both of you use samples where you have done some racial selection, but is it possible in a country like the States where you have more and more mixed couples and mixed identities to have a purely "white" sample? Seems to me quasi impossible now. So, how do you do that? How do you select the sample?
The second question is about gender? Do genetic factors work exactly the same for men and women? I saw in some of your regressions nice odds ratios for gender impact, could you say a little more?
Last, what about ethics? What are the constraints, the ethical rules for that type of study? For the small scale ethnographic research we do, we have a long list of what we have and haven't the right want to do. How do you manage to get authorization to do that type of survey? What are the rules to follow?

Franck Ramus

I'll start by the end. With respect to ethics, this is controlled from the very beginning of the study.
When the study is designed, it has to be submitted and approved by an ethics committee. When the participants are recruited, they have to be informed about the

purposes of the study and they have to sign an informed consent form, which delimitates the possible uses of the data that they provide.

Some studies, for example, the UK Biobank, are limited to medical questions. This was part of the consent form, and therefore you're not supposed to use this data to venture too far away from medical questions that could potentially contribute to improving health.
In other studies, they didn't have such limitations and the participants agreed for their DNA to be used in relation to any other data that is available.
In all cases, researchers like us who come after the initiation of the study just benefit from this consent that was given by participants.

The topic of gender is indeed fascinating. Males and females are differentiated by their sex chromosomes, but they are often not analysed in the Genome-Wide Association Studies that we have presented. However, for those intrigued by phenotypes that vary between males and females, this becomes a subject of study in its own right.

The focus is not solely on the impact of genetic variations on sex chromosomes, but also on the extent to which the effects of genes on autosomes (all the other chromosomes that are the same in males and females) might differ between the two sexes. And also on the extent to which the same genes may be expressed differently, in different quantities in different tissues, between males and females. Indeed, there are projects underway that aim to elucidate the genetic basis of some gender differences. This is a highly intriguing field, albeit one that is fraught with controversy and sensitivity.

Nonna Mayer

Thank you. Have you published anything on the topic?

Franck Ramus

I haven’t. But one of my former students has a project on this topic, and will start it when she comes back to France.

Last, the question on mixed ancestry.
For the moment, genetic studies strive to carry out their analyses based on relatively homogeneous samples, excluding participants from populations that are less well represented and those with highly mixed ancestry. But of course, we all know that ancestry does not come in discrete categories, and that most people have to some extent mixed ancestry. Ancestry is actually a notion that is best represented on a continuous multidimensional space, with each individual one dot in that space.
I imagine that at some point, when we have enough GWAS data on all of the world's population and a better understanding of genetic diversity, the community will move on from studies restricted to homogeneous populations to studies including the full diversity of human populations, but statistically adjusting on multiple ancestry components, in order to avoid the gene-culture confounds that we mentioned earlier.

Nonna Mayer stops the discussion as another seminar is going to start in the same room. Shewarmly thanks the two speakers for making clearer the complex issue of genomics, and to all the participants whose questions fueled the debate.