Ayyar, Sreevidya; Bolt, Uta; French, Eric; O'Dea, Cormac

Working Paper

Imagine your life at 25: Gender conformity and later-life outcomes

IFS Working Papers, No. 24/17

Provided in Cooperation with:
The Institute for Fiscal Studies (IFS), London

Suggested Citation: Ayyar, Sreevidya; Bolt, Uta; French, Eric; O'Dea, Cormac (2024) : Imagine your life at 25: Gender conformity and later-life outcomes, IFS Working Papers, No. 24/17, The Institute for Fiscal Studies (IFS), London, https://doi.org/10.1920/wp.ifs.2024.1724

This Version is available at:
https://hdl.handle.net/10419/324400

Institute for Fiscal Studies

Sreevidya Ayyar
Uta Bolt
Eric French
Cormac O'Dea

Working paper
24/17

Imagine your life at 25: Gender conformity and later-life outcomes

Economic and Social Research Council

Imagine your Life at 25: Gender Conformity and Later-Life Outcomes

Sreevidya Ayyar* Uta Bolt† Eric French‡ Cormac O'Dea§¶

April 24, 2024

Abstract

Using a digitized sample of thousands of essays written by 11-year-olds in 1969, we construct an index which measures the extent to which girls' imagined futures conform to gender norms in Britain at the time. We link this index to outcomes over the life-cycle. Conditional on a large set of age-11 covariates, a one standard deviation increase in our index is associated with a decrease in lifetime earnings of 3.5%, due to both lower wages and fewer hours worked. Half of this earnings decline is mediated by reduced educational attainment, selection into lower-paid occupations, and earlier family formation of those who conform more strongly to prevalent gender norms. Holding skills constant, girls whose essays conform less to gender norms, live in regions with higher female employment and educational attainment. This highlights that the wider environment in which girls grow up shapes gender conformity.

Keywords: Gender, Children, Text Analysis

JEL codes: J16, J13, Z13

*London School of Economics and Political Science

†University of Bristol and Institute for Fiscal Studies

‡University of Cambridge and Institute for Fiscal Studies

§Yale University, NBER, and Institute for Fiscal Studies

¶We are grateful to Richard Gong for excellent research assistance and to Lila Alloula, Loui Chen, Patrick Chu, Jun-Davinci Choi, Aaron Cullen, Mitchell Dubin, Alexandra Gers, Alessa Kim-Panero, Rosa Kleinman, Ethan Liu, Nastaran Moghimi, Eleri Phillips, Charlotte Townley, and Teddy Tawil for correcting spelling mistakes in the essays. Thanks to Barbara Biasi, Enda Hargaden, seminar participants at Bonn, Cambridge, LSE, Manchester, University College Dublin, the IFS Conference Intergenerational Inequality, and the Zurich Text-As-Data conference for comments. Funding – the Economic and Social Research Council ("Centre for Microeconomic Analysis of Public Policy at the Institute for Fiscal Studies" (RES-544-28-50001), "Intergenerational transfers, insurance, and the transmission of inequality" (ES/V001248/1), and NORFACE for Grant 462-16-120: Trends in Inequality: Sources and Policy" – for this work is gratefully acknowledged. We are grateful to Centre for Longitudinal Studies (CLS) for use of these data and to UK Data Service for making them available. Opinions and conclusions are solely those of the authors. Errors are our own.

1 Introduction

Despite a narrowing gender earnings gap, considerable differences in earnings between men and women still exist (Goldin, 2021). Gender norms, and in particular the extent to which women conform to these prevailing norms, can potentially play a substantive role in explaining such earnings differences (Fortin, 2005; Bertrand, 2020).[1]

In this paper, we use novel text data – over ten thousand essays written by 11-year-olds in 1969 – to construct an index of gender conformity, which measures the extent to which girls imagine their future to conform to gender norms at the time. We find that, conditional on a rich set of age-11 characteristics including family background variables, as well as cognitive and non-cognitive skills, girls who exhibit one standard deviation stronger gender conformity in their essays have 3.5% lower lifetime earnings. Half of this association can be explained by a combination of lower educational attainment, selection into lower-paid occupations, and earlier family formation of girls who display stronger gender conformity. How strongly a girl conforms to gender norms is associated with her skills, and the characteristics of the region she grew up in; those growing up in areas with higher female employment and university attendance rates tend to exhibit less conformity with prevailing gender norms in their essays.

Our data come from the National Child Development Study (NCDS), which is an ongoing panel survey containing almost the entire population of Britain born in a particular week in 1958. At age 11, all respondents were asked to write an essay, during school hours, about the life they imagined having at age 25. These essays have recently been digitized by the Centre for Longitudinal Studies (CLS). The occurrence of certain topics in these essays has been shown to be predictive of later life outcomes such as health, cognition, and income (Pongiglione et al., 2020; Wielgoszewska et al., 2022; Kern et al., 2022).

In this paper, we extract a measure of gender conformity from girls' essays using natural language processing. Our key identifying assumption is that girls conform more strongly to prevalent gender norms if the content and style of their essay is typically associated with females. For example, describing family and housework (as opposed to career and football) is more typically feminine. We remain agnostic about the reasons behind gender conformity – girls may imagine their futures to be more aligned with female gender norms due to (perceived) constraints or due to different preferences. We discuss how each of these could

[1] Following Brenøe (2022), we define gender norms as society's perceptions of how women and men should behave in general within society. Gender conformity is the act of adhering, through one's own actions and behaviors, to the prevailing gender norms in society. Whilst gender norms in a given society tend to be fixed at a given point in time, the degree to which individuals conform to those norms differ.

rationalize our findings using a model of education choice and time allocation in Section 5.

To estimate a measure of gender conformity, we first manually correct over 100,000 spelling mistakes in the essay data. Applying a method proposed by Kozlowski et al. (2019), we then train a Word-Embedding Model using 1 million books written between 1958 and 1978 to measure how "feminine" or "masculine" individual words were at the time the essays were written. Words which frequently appear in context with unambiguously gendered words like "women", "girl", "she", "her" etc., are adjudicated by the algorithm to be more feminine. Using measures of how feminine individual words are, we capture the strength of female associations in each essay, which in turn gives us a measure of the writer's gender conformity.

Holding age-11 measured skills and other characteristics constant, girls with one standard deviation stronger gender conformity have 3.5% lower lifetime earnings. This is due to both lower wages and, to a lesser extent, fewer hours of work. To uncover *why* girls with stronger gender conformity are predicted to have lower lifetime earnings, we perform mediation analysis. This allows us to estimate how much of the link between lifetime earnings and gender conformity is explained by gender conformity affecting educational attainment, occupational choices, and family formation decisions. A one standard deviation increase in our index of gender conformity predicts a 1.1 percentage point lower probability of attaining a university degree and a 1.6 percentage point decrease in the probability of entering a professional occupation. Girls who conform more strongly to gender norms also tend to get married earlier and have children earlier. Overall, just over half of the association between lifetime earnings and gender conformity is mediated through education, occupation, and family formation. We then turn to the question of what determines gender conformity. We find that girls with better skills and those who grew up in regions with higher female employment and educational attainment display less conformity with prevalent gender norms.

Our paper is among the first to study the extent to which young girls internalize prevailing gender norms, as captured by whether they prefer (or expect) to conform to these norms in their own future lives. This is in contrast to three existing approaches in the literature on measuring the association between gender norms and outcomes. The first is to proxy a woman's gender conformity and attitudes using gender norms prevalent in her country of origin, since these norms have been found to persist across generations (Alesina et al., 2013). Prominent examples of this approach are Fernández and Fogli (2009) and Boelmann et al. (2021), who use data on migrants, to the US and within Germany respectively, to study how gender norms towards work and fertility in a woman's country of origin influence her labor market decisions. Both papers find that such norms have significant explanatory power for

a woman's labor supply, even after she moves to a different country with different economic and institutional conditions. A second prominent approach in the literature is to directly measure attitudes towards, and conformity with, gender norms. For example, Vella (1998), Fortin (2005), and Cavapozzi et al. (2021) all use survey questions which ask individuals the extent to which they agree with a set of statements about the role of women and men in society. They find a strong negative relationship between support for traditional gender norms and women's labor supply. Field et al. (2021) measure survey respondents' preferences and beliefs about whether women should work outside the home (in addition to their perceptions of community norms) and show how experimentally shifting those beliefs can increase female labor supply. Banan et al. (2023) study gender non-conformity among young girls and boys by investigating differences in their stated preferences over hobbies, time use, and school subjects and linking these measure to outcomes in adulthood. A third group of papers focuses on more indirect evidence of the existence of gender norms and conformity to them. For instance, Bertrand et al. (2015) find that the density of a wife's share of earned income exhibits a sharp drop when it exceeds 0.5, among a sample of married couples in the US. This is consistent with the view that gender norms induce an aversion to the wife earning more than her husband. Wiswall and Zafar (2018) find that females have stronger preferences for certain workplace amenities using surveys of undergraduate students. These preferences can explain 25% of the early-career gender wage gap.

Our approach enjoys three distinct advantages. First, because respondents were asked to write about their future and not prompted to consider gender or answer any questions related to gender, our approach avoids their being primed. This alleviates concerns about social desirability bias or experimenter demand effects (Bertrand and Mullainathan, 2001; De Quidt et al., 2019). Second, we measure gender conformity at a young age, *before* career and family decisions are made. Our study is thus less susceptible to reverse causality or justification bias (Black et al., 2017). The third advantage is that we can link our measure of gender conformity not only to outcomes through different stages of adulthood, but also to skills and circumstances in early childhood. Key recent studies show that attitudes towards gender norms are malleable through discussions and information (Dhar et al., 2022; Bursztyn et al., 2020). Evidence also exists that family background characteristics, such as sibling gender composition affect conformity with gender norms (Brenøe, 2022).

By studying the language of children at a large scale, we add to a recent literature in economics which uses text as data. Previously, sources of texts have ranged over political speeches (Gentzkow et al., 2019), newspaper articles (Cagé et al., 2020), financial disclosures

(Hanley and Hoberg, 2019), patent filings (Kelly et al., 2021), online message boards (Wu, 2020), transcripts from policymaker committees (Hansen et al., 2018), and college syllabi (Biasi and Ma, 2022). Our work is a novel application of a methodology originally proposed in Kozlowski et al. (2019), who use texts written between 1900 to 1999 to study how the various dimensions of social class (gender, affluence, education etc.) have evolved over the twentieth century. Ash et al. (2024) build on this same methodology and using published opinions from circuit courts in thr US to obtain measure of judges' gender bias.

Two related studies which use natural language processing are Adukia et al. (2023), who study the race and gender representations contained in children's books, and Michalopoulos and Xue (2021), who study folklore tales.[2] Our paper complements both of these studies in that we show how children, the main consumers of such texts, have already internalized gender-based norms at a young age, and in a way which matters for their lifetime outcomes.

Our paper proceeds as follows. Section 2 describes the NCDS data. Section 3 discusses how we estimate gender conformity. Section 4 contains our results, while Section 5 presents a simple model which can rationalize our results. Section 6 concludes.

2 Data

Our data comes from the National Child Development Study (NCDS).[3] The data covers almost all children born in Great Britain in the third week of March 1958 and follows them to this day. It is unique in that it provides information on skills and development in early childhood as well as on work, family, and earnings over the working life in adulthood.

The initial survey at birth has been followed by subsequent surveys at ages: 7, 11, 16, 23, 33, 42, 46, 50, and 55.[4] The data from childhood includes measures of cognitive and non-cognitive skills, family circumstances, and parental income. Later waves of the study record educational outcomes, family formation (child-bearing, marriage, divorce etc.), earnings, and hours of work. As part of the age 11 wave, respondents wrote the essays that are central to our study (Centre for Longitudinal Studies, 2018).

[2]Whilst Adukia et al. (2023) apply natural language processing (NLP) directly to books, Michalopoulos and Xue (2021) use the NLP-based ConceptNet to categorize a list of motifs.

[3]The National Child Development Study is provided by the Centre for Longitudinal Studies (2023a) at the Institute of Education, University College London.

[4]We do not use the information from the age 46 survey, as it was a limited telephone-based survey.

2.1 Sample Selection

The NCDS panel data contains 10,511 essays, among which 5,091 were written by girls. For the purpose of estimating our index of gender conformity, we use all 10,511 essays.[5]

Our analysis sample is formed of girls who wrote essays for whom we observe earnings, wages, employment status, and hours worked at least once between the ages of 23 and 55. The population of the UK born in 1958 was overwhelmingly white, and we drop the small number of non-white girls.[6] The final sample for our analysis consists of 3,497 girls. For more details, see Table B.1.

2.2 The Essays

During the child interview component of the age 11 (1969) survey wave, NCDS cohort members were asked to write an essay on what they imagined their lives would be like at 25. They were given 30 minutes to write the essay. The instruction was:

> *Imagine you are now 25 years old. Write about the life you are leading, your interests, your home life and your work at the age of 25.*

Table 1 provides summary statistics on the full sample of 10,511 essays, the sample of 5,091 essays written by girls, and our analysis sample of 3,497 essays. For our analysis (full) sample, mean essay length is 230 (201) words, of which on average 10.3 (11.3) words are misspelled. The differences between the analysis sample and the full sample largely reflect the better writing abilities of girls in this age group, relative to boys; indeed, comparing the second and third panels of Table 1 reveals that our analysis sample is representative of all essays written by girls. Figure 1 depicts the most frequent words across our analysis sample (excluding articles and pronouns) – this illustrates the diverse content in the essays, which ranges from sport and domestic chores to career choice and family life.

[5] Goodman et al. (2017) describe that 90% of cohort members who participated in the age 11 wave wrote essays and that the characteristics of the essay writers are largely representative of the population.

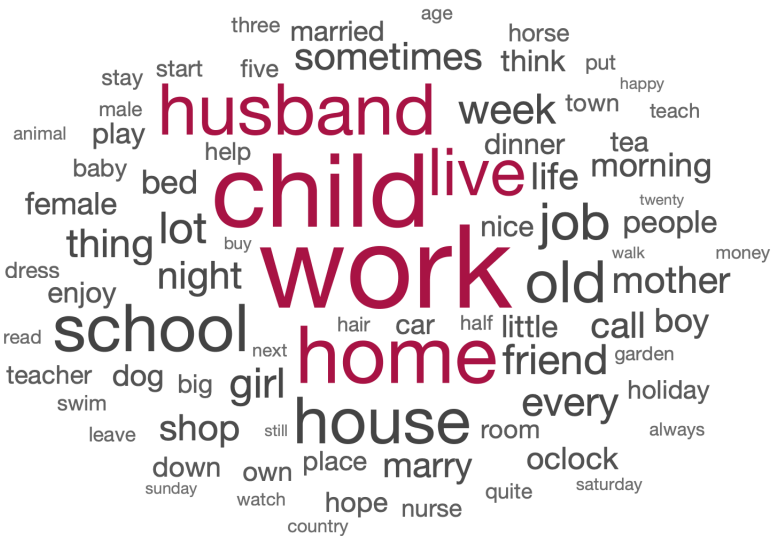
[6] The sample of non-white girls for whom we observe essays and outcomes is small (48 individuals) and our results are not sensitive to their exclusion. Preliminary analysis showed that non-white girls are less gender conforming than white girls, though we consider the sample too small to draw conclusions.

Table 1: Describing the Essays

	Mean	Standard Deviation	25th Percentile	50th Percentile	75th Percentile
Full Sample					
Essay Length (words)	201.0	107.1	125	185	259
Spelling Mistakes	11.3	9.6	5	9	15
Girls Sample					
Essay Length (words)	225.9	113.8	145	210	288
Spelling Mistakes	10.6	9.0	5	8	14
Analysis Sample					
Essay Length (words)	230.7	113.5	149	216	292
Spelling Mistakes	10.3	8.7	5	8	14

Notes: The full sample is 10,511 essays, which is the total number of digitized essays in the NCDS data. The girls sample refers to (all) essays written by girls (5,091) and the analysis sample is 3,497 essays.

Figure 1: Eighty Most Frequent Words



Notes: The graphic shows the 80 most frequently used words in our analysis sample (3,497 essays). The highlighted words are the five most-used, and words which occur more often have a larger font size.

To motivate why these essays might be informative about gender conformity, we obtained permission to reproduce excerpts from two (uncorrected) essays.[7]

Essay 1: "*There was a pile of washing waiting to be done in the laundry basket waiting for me to wash. Oh how I wish I was young I thought to myself. Just as I had the water in the washer I heard my five month old baby crying in her pram outside. It was her bottle time I have to leave every thing to get her bottle ready. As soon a I had fed her my husband came home for his dinner. It had to be a ham (sand/samwhich) samwidgch today so I could get my washing done quicker. [...]*"

Essay 2: "*If I was 25 now, I would probably live in a small flat. I would go to work at 9 o'clock in the morning and finnish at 5 o'clock, I would make my own meals. After my tea I would go out in my car. When summer came round I would go away for a week say Paris or New York. My flat would have modern furniture in it. I would hold lots of parties and have beer, sausage on sticks, salad and Chinese food. [...]*"

A casual read of the essays suggests that the content of essay 1 conforms more with traditional gender norms regarding work and home production relative to the content of essay 2. In Section 3 we formalize this notion by constructing a measure of gender conformity.

2.3 Childhood Data

At ages 7, 11, and 16, we observe comprehensive data on the NCDS cohort members. These include: several measures of cognitive and non-cognitive skills, county of residence, and family background variables such as parental education, parental income, sibling composition and family stability. We discuss each of these variables below.

Cognitive Skills As part of the age 11 survey, cohort members took part in math, reading, and drawing tests. We interpret the resulting test scores as noisy measures of unobserved, underlying cognitive skill. Following previous literature (Heckman et al., 2013), we use a latent factor model to combine these measures to a cognitive skill index which minimizes measurement error (see Appendix B.2 for details). We standardize the cognitive skills index to have mean zero and variance of 1. Higher values of the index reflect higher cognitive skills.

Non-Cognitive Skills Several papers have highlighted the importance of non-cognitive skills in childhood for explaining lifetime outcomes (Heckman et al., 2006; Papageorge et al., 2019; Todd and Zhang, 2020). Non-cognitive skills capture a child's social adjustment and temperament. We have several teacher- and parent-reported measures of cohort members' behaviors at age 11, which we consider noisy measures of latent non-cognitive skills. We

[7]We are grateful to the Centre for Longitudinal Studies Data Access Committee for permission to reproduce these excerpts. All essay files can be downloaded from the UK data archive (https://www.data-archive.ac.uk/) on acceptance of the terms of use.

use the same latent variables approach as we do for cognitive skills to construct an index of non-cognitive skills. Details on this procedure are also in Appendix B.2. A higher value of the index indicates a "better-behaved" child, with higher non-cognitive skills.[8]

Family Background Variables Comprehensive data on parental income was collected when the NCDS cohort members were 16. Parental income is the sum of father's earnings, mother's earnings, and other income (net of taxes). More details are given in Appendix B.3.

We have information on years of schooling completed by each parent of an NCDS sample member. Furthermore, in the age 11 survey, there are variables detailing a cohort member's birth order, and number and sex of siblings. The same survey wave contains information on the educational aspirations that parents have for their children, measured by a survey item which asks whether the parents hope their child will pursue further education. Finally, we observe whether the cohort member's parents are divorced or separated.

Geographic Variables In addition to individual-level covariates on skills, behaviors, and family background, the survey records region of residence (of which there are 11) at age 11. We use this data to estimate region-level female employment and university-attendance rates using information on the mothers of the NCDS cohort – we construct a leave-one-out measure for each cohort member, excluding the mother of the cohort member of interest. The survey also contains location data at a finer level of aggregation in the subsequent survey wave (at age 16), at which point county of residence (of which there are 111) is collected. In our main empirical specification we include county fixed effects, where we set a cohort member's county to be missing if they moved region between age 11 and age 16.

2.4 Adulthood Data

Outcomes in adulthood are constructed from surveys at ages 23, 33, 42, 50, and 55.

Earnings We observe gross weekly earnings in every survey wave, which we convert to a monthly measure. We impute monthly earnings for the period between interviews by first regressing earnings on an individual fixed effect and educational attainment interacted with a time fixed-effect. Using the NCDS activity histories, we set earnings in a month equal to zero if an individual is recorded as not being employed in that month. Our measure of lifetime earnings is the (undiscounted) sum of all real monthly earnings between 23 and 55.

[8]More specifically, measures which record whether the cohort member is unconcentrated, fidgety, miserable, or destructive correlate negatively with the index, while measures which record perceived obedience and socialising skill correlate positively.

Wages We define a measure of hourly wages by dividing weekly earnings by weekly hours of work. We construct hourly wages at ages 23, 33, 42, 50, and 55. Using the same fixed-effects procedure as for earnings, we impute hourly wages for non-interview months across 32 years. If a cohort member is not employed in a given month, we set their wage to missing. Average hourly wages (across all working months) over the life cycle are the simple mean of hourly wages between 23 and 55 inclusive (observed and imputed).

More information on how we construct lifetime earnings and wages is in Appendix B.

Labor Supply The NCDS activity history data provides a detailed record of employment status and type of work (part- or full-time) for each month between the age 23 and age 55 surveys (inclusive). We compute share of time spent in employment by finding the fraction of months cohort members spent either in full- or part-time employment. Total hours worked over the life cycle sums across monthly hours worked (which are computed based on 40 hours/week for full-time work and 20 hours/week for part-time work and 0 for those not working). Our measure of total hours is *not* computed conditional on employment.

Educational Attainment Cohort members reported their highest level of educational qualification at age 33. We use this to define a categorical measure of education, distinguishing between those who have compulsory education only (qualifications lower than O-levels), those who have some post-compulsory education (leaving education with O-or A-levels), and those who have a degree higher than A-levels.

Family Formation We observe whether a cohort member is married, as well as how many children they have, at ages 23, 33, 42, 50, and 55.

Occupation For an NCDS cohort member's occupation, we use the age 33 occupational classifications constructed by Gregg (2012), which in turn are based on the Standard Occupational Classification 2000 (SOC2000) by the Office of National Statistics in the UK. Occupations are classified into four categories, based on the standard coding of social class: professional, administrative (non-manual and skilled), manual, and elementary (unskilled).

3 An Index of Gender Conformity

3.1 Identifying Gender Conformity

To measure gender conformity, the key identifying assumption we make is that girls with stronger gender conformity write essays whose content and style is typically associated with femininity. For example, describing family and housework would be considered more typically

feminine than writing about work and football. Using reflective words ("feel" or "wish") instead of factual language is also associated with being female (Argamon et al., 2003).

The constructive step in applying our identifying assumption is how one defines which roles, norms, and activities are typically associated with being feminine. In general, this will differ across countries and time. For instance, Kozlowski et al. (2019) provide evidence that the association of gender with affluence, education, and status has shifted over the 20th century. Akerlof and Kranton (2000) and Benjamin et al. (2010) emphasize the importance of measuring gender identity relative to societal beliefs, since people adopt identities that conform with socially-prescribed behavior. Due to this, we deem it important to identify the gender associations of the essays in the context in which they were written (in Britain in 1969). Hence, we evaluate how feminine the content of an essay is, and therefore measure gender conformity, relative to texts written in Britain between 1958 and 1978.

The following two points are noteworthy for interpreting what we measure. First, we estimate the extent to which girls' imagined futures conform to gender norms at the time. We take this as a measure of their gender conformity when it comes to labor market relevant decisions and label this "gender conformity" for brevity.[9] Secondly, our gender conformity index captures how much girls' attitudes towards their *own* future are in line with traditional norms; it does *not* capture what they think other women or women in general should do.

In the remainder of this section, we outline how we measure gender conformity. We first give an overview of our procedure in Section 3.2, before detailing each step in Section 3.3. Section 3.4 contains descriptive information about our index, as well as an application of our procedure to the two example essays presented previously.

3.2 A Roadmap for Estimating Gender Conformity

Our estimation procedure has six steps. The first is to spell-correct the more than 100,000 spelling mistakes across the full sample of 10,511 essays. Misspelled words were automatically flagged if they were not found in a UK English Dictionary. A team of undergraduate research assistants read these flagged words in the context of the sentence and corrected them manually. The second step is to represent each essay as a set of words, with associated word counts. The third step is to assign a vector value to each word in the essays. These vector values encode a word's meaning and association, such that words with similar meaning or words that are used in similar contexts are given similar vector values. The fourth step is to construct a "gender dimension" by averaging across the difference in vector values

[9] Another approach is Banan et al. (2023), who measure gender conformity using information on the gender composition of a child's friendship group and whether their current hobbies align with gender norms.

(from step three) of pairs of unambiguously gendered words (e.g. woman−man, girl−boy, female−male). The fifth step is to determine how feminine or masculine each word in the essay is by projecting its vector value onto the gender dimension constructed in step four. In the sixth and final step, we construct a weighted sum of all words for each essay, where a weight is given by a word's projection value. After residualizing this weighted sum on a quadratic of essay length, and standardizing residuals to have mean zero and variance of one, we arrive at our "index of gender conformity".

We now explain this procedure in more detail.

3.3 Estimating Gender Conformity

3.3.1 Step 1: Digitisation, spelling correction, and pre-processing

The original NCDS essays from 1969 were handwritten. In 2018, the Centre for Longitudinal Studies (CLS) digitized the essays, redacting only names and other potentially identifying information from each essay.

To make our data usable for standard text processing software, we correct spelling mistakes in the essays. First, we identify misspelled words using the UK English dictionary in Python's *pyenchant* library. A team of undergraduate research assistants then manually correct any words that are identified as misspelled. Correcting misspelled words individually allows us to maintain the essay's original structure and message. Whilst automatic spelling correction is available, it is ill-suited to the phonetic nature of children's spelling mistakes (Downs et al., 2020). As examples, note the use of "hoosban" for husband or "aspesaly" for "especially". Because these spellings are based on the phonetic pronunciation of words, instead of being typos which are one or two letters away from the correct spelling, most standard spell-check software would not correctly replace these misspelled words.[10] Thus, we do so manually. Upon making spelling corrections, we arrive at our final essay data.

3.3.2 Step 2: Representation of essays as counts of words

In all that follows, we will treat each essay as a matrix of word/phrase counts. This is known as a *Bag of Words* representation of text data. First, we remove punctuation and stop words from the essays (such as articles and pronouns). Then, we count all individual words (also known as 1-grams). We also count all 2-word phrases (or 2-grams) in the essays, and we add to our matrix of counts those 2-grams whose first word is "not" or "no". This is so that we

[10]More powerful spell-check software, such as the one used in Microsoft Word, often offer the correctly-spelled word as one among a menu of options, but rarely as the first choice or "best-guess". Manual selection of the correct spelling would still be required.

can distinguish, for example, between references to being "married" versus "not married".

Table 2: Example - Bag of Words

friend	husband	married	not married
2	1	1	1

Table 2 illustrates what our Bag of Words procedure would look like for the following sentences: "My friend is not married. Does your friend have a husband?" Our procedure first drops the stop words: "my, is, does, your, have, a". It then counts 1-grams: "friend, married, husband", and adds to the matrix counts for the 2-gram "not married". All other 2-grams, such as "friend married", are ignored. Our representation of essays is similar to the representation of these two sentences, with over 18,000 words and associated word counts.

3.3.3 Step 3: Assign vector value to words using a Word-Embedding Model

Word-Embedding Models (WEMs) uniquely represent words as high-dimensional, real-valued vectors. Words with similar meanings and/or association receive similar vector values. Once words are assigned vector values, relationships between words can be quantified using linear algebra. We train a WEM to assign words in our Bag of Words matrix to a numerical vector.

While off-the-shelf WEMs are available, we are not aware of any that are specific to the time period around when the essays were written (in 1969). Because word meanings and associations potentially change over time, we choose to construct our own.

We train Google's skip-gram Word2Vec algorithm[11] whose purpose is to predict, for each word, adjacent words in a sentence (Mikolov et al., 2013). For example, if we input "girl", the algorithm might predict "the girl put on a blue dress". To make such predictions, this algorithm performs unsupervised learning. That is, it "learns" about text patterns from data used to train it, such as how frequently pairs of words appear together, how far apart words occur, and whether some words always follow others. Based on what it "learns", the skip-gram Word2Vec algorithm assigns each word a numerical vector value; words which predict similar adjacent words are assigned similar vector values (Sabra and Sabeeh, 2020). The Word-Embedding Model provides a mapping from the words to the real-valued vector.

The data we use to train our skip-gram Word2Vec algorithm is the Google Books N-grams Version 3 data; specifically, we use the English 5-grams data.[12] We use all 1-,2-,3-,4-

[11] We choose the skip-gram class of algorithms because they have been shown to perform best when trained on a lot of text data which contains words that are used infrequently (Sabra and Sabeeh, 2020).

[12] storage.googleapis.com/books/ngrams/books/datasetsv3.html

and 5-grams[13] from over 2.6 million books/documents written between 1958 and 1978.[14]

The only aspect of the training procedure that we choose is the dimensionality of the vector values that are assigned to words. There is a trade-off involved in this choice. The higher the dimensionality we allow for, the more nuanced the word relationships encoded by the algorithm, but the lower the precision of the vector values assigned to each word. We specify that each word in the Google Books N-grams data be represented by a (300×1) vector, following prior research which finds that a 300 dimensional vector allows for nuance in words' meanings but still provides precision (Patel and Bhattacharyya, 2017).

Although the 300 elements of the vector value for each word assigned by the Word2Vec algorithm are difficult to interpret, the algorithm does construct representations of words with intuitive structures. For example, consider the words *king*, *man*, *woman*, *queen*, each of which is associated with the (300×1) vector values $\vec{king}$, $\vec{man}$, $\vec{woman}$, $\vec{queen}$. It turns out that our Word-Embedding Model produces the following insight: $\vec{king} - \vec{man} + \vec{woman} \approx \vec{queen}$. These sorts of analogies are common to WEMs (Pennington et al., 2014). We exploit this feature of WEMs to measure the gender association of words in step five.

3.3.4 Step 4: Constructing a Gender Dimension

Kozlowski et al. (2019) argues that the reason that $\vec{king} + (\vec{woman} - \vec{man}) \approx \vec{queen}$ is that $(\vec{woman} - \vec{man})$ acts as a "gender dimension". Adding $(\vec{woman} - \vec{man})$ to $\vec{king}$ has the effect of starting at $\vec{king}$, and taking one step in the direction of femininity. Not only is $\vec{king} + (\vec{woman} - \vec{man}) \approx \vec{queen}$, but $\vec{king} + (\vec{female} - \vec{male})$ also approximately equals $\vec{queen}$. This suggests that the "gender dimension" is captured not only by $(\vec{woman} - \vec{man})$, but also by $(\vec{female} - \vec{male})$, $(\vec{girl} - \vec{boy})$, $(\vec{her} - \vec{him})$ etc. More generally, an approximation of Kozlowski et al. (2019)'s "gender dimension" can use any pair(s) of words whose semantic difference corresponds to the feminine-masculine distinction.

Using these insights, we can construct a *gender dimension vector*, $\vec{GD}$, in our word-

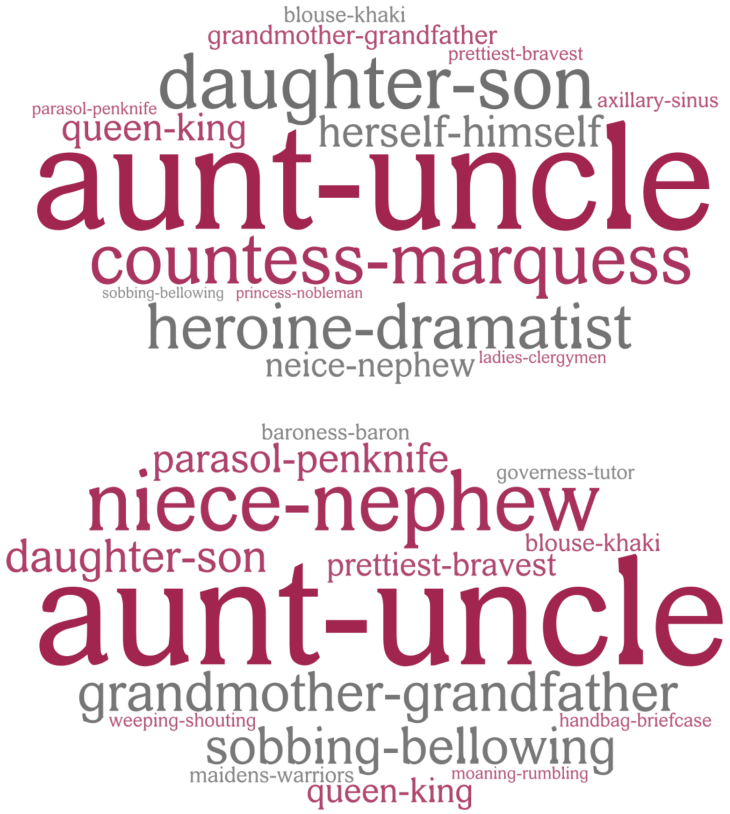
[13] 1-,2-,3-,4- and 5-grams are used to train the Word2Vec algorithm. Including longer phrases in the training data improves the algorithm's ability to detect text patterns. Since the algorithm ultimately assigns vector-values to individual words, we restrict our attention to 1-grams and select 2-grams from the essays.

[14] As a robustness check, we re-run our analysis with two pre-trained WEMs. The first is from Google that is part of an open-source project (code.google.com/archive/p/word2vec). Google trained a Word2Vec algorithm using primarily Google News data, which contains about 100 billion words. The second WEM is trained by Pennington et al. (2014) using a different algorithm known as GloVe.

embedding model by taking the average of the vector values of 10 pairs of gender words

$$\begin{aligned}\vec{GD} \equiv \frac{1}{10} \quad \{&(\vec{woman} - \vec{man}) \quad +(\vec{women} - \vec{men}) + (\vec{she} - \vec{he}) + (\vec{her} - \vec{his}) \\ &+(\vec{her} - \vec{him}) \quad +(\vec{hers} - \vec{his}) + (\vec{girl} - \vec{boy}) + (\vec{girls} - \vec{boys}) \\ &+(\vec{female} - \vec{male}) \quad +(\vec{feminine} - \vec{masculine})\}.\end{aligned}$$

Figure 2: Pairs of Gendered Words



Notes: These word clouds each plot 15 pairs of words which lie closest to the gender dimension $\vec{GD}$. That is, the pairs (a, b) which maximise $cos(\vec{a} - \vec{b}, \vec{GD})$. The top figure uses words from the entire WEM (of which there are over 2 million), while the bottom one is restricted to pairs of words found in the essays (approx. 18,000). A larger font-size indicates that a word-pair is closer to $\vec{GD}$. Colors are *not* meaningful.

Showing specific analogies like $\vec{king} - \vec{man} + \vec{woman} \approx \vec{queen}$ may not be sufficient evidence that a WEM is able to capture gender associations between words, and therefore that our gender dimension is meaningful (Nissim et al., 2020; Zhang et al., 2020). To address such concerns, we implement Bolukbasi et al. (2016)'s recommendation of checking whether, when starting from our gender dimension $\vec{GD}$, our Word-Embedding Model can identify pairs of gendered words such as $\vec{king} - \vec{queen}$. To do this we use Bolukbasi et al. (2016)'s algorithm

to select the fifteen word pairs whose differenced vector values lie closest to $\vec{GD}$ (as measured by their cosine similarity). We first apply the algorithm to all possible word pairs in our Word-Embedding Model and second to all word pairs in the essays. Figure 2 illustrates our results. The fact that these pairs are explicitly gendered, and even include $(\vec{king} - \vec{queen})$, validates that $\vec{GD}$ does indeed capture the feminine-masculine distinction.

3.3.5 Step 5: Determining the gender associations of words

Having shown that our WEM captures gender associations between words, we now turn to measuring the gender association of words in our essays. Using the vector values we assigned to each word in step two, we project each word's vector onto the gender dimension vector, $\vec{GD}$, by calculating the cosine similarity, which is a common way of measuring how similar words (or in this case a word and $\vec{GD}$) are in vector space (Sabra and Sabeeh, 2020).

For example, to compute the gender association of "married", we project $\vec{married}$ onto $\vec{GD}$ by finding their scalar dot product:

$$\vec{married}^T \cdot \vec{GD} = \begin{pmatrix} married_1 & married_2 & \dots & married_{300} \end{pmatrix} \cdot \begin{pmatrix} GD_1 \\ GD_2 \\ \vdots \\ GD_{300} \end{pmatrix}$$

$$= married_1 GD_1 + married_2 GD_2 + \ldots + married_{300} GD_{300}$$

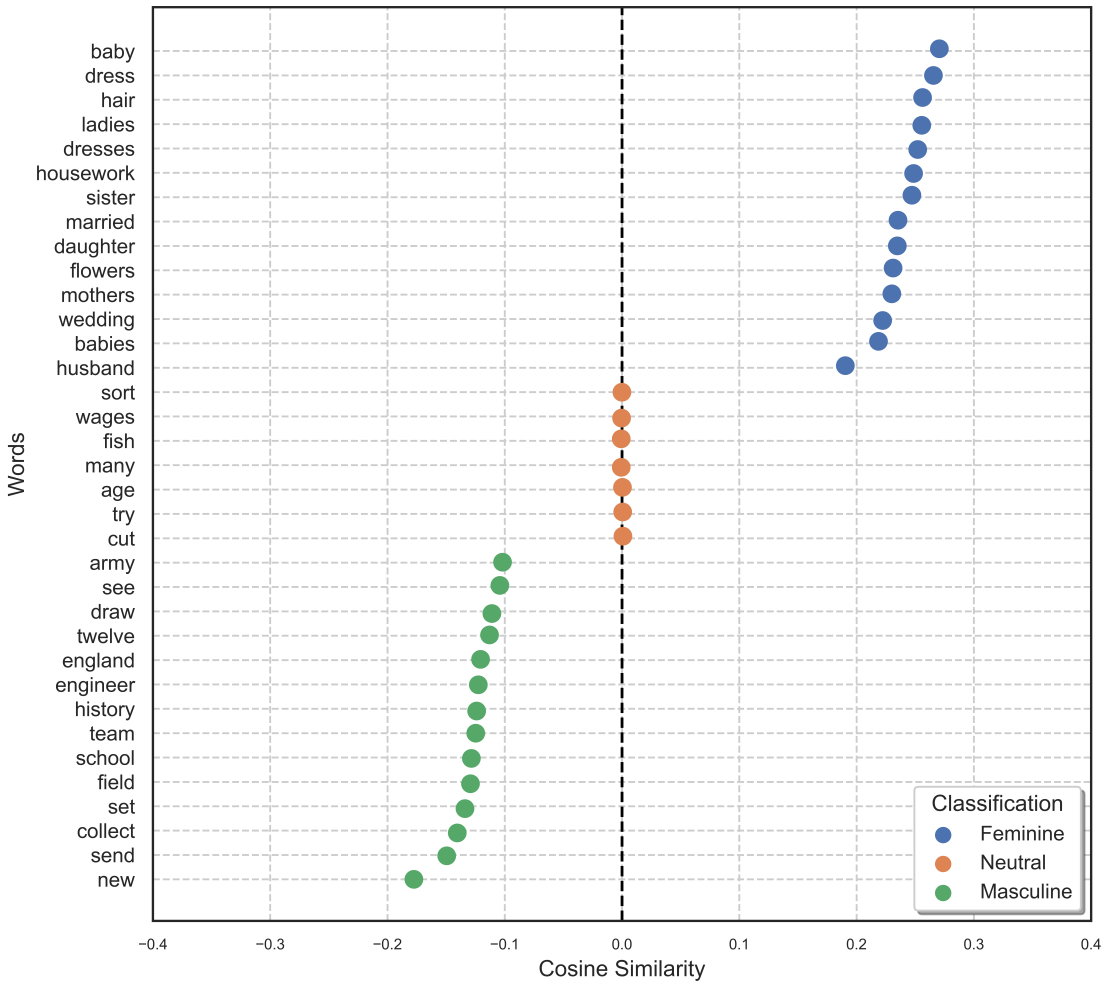
where $married_i$ and GD_i are the values of the i^{th} elements of the $\vec{married}$ and $\vec{GD}$ vectors respectively. A positive projection value (between 0 and 1) would imply "married" is associated with being feminine, while a negative projection value (between -1 and 0) would indicate a masculine association. The magnitude of the projection value indicates how gendered a word is. The larger the absolute value of the projection value, the more strongly masculine or feminine the word is. Using our WEM, we find $\vec{married}^T \cdot \vec{GD} = 0.23$.

Figure 3 plots projections of: (i) the 15 most feminine words (with the most positive projection values), (ii) the 15 most masculine words (with the most negative projection values) and (iii) 10 neutral words (such that their projection onto $\vec{GD}$ is zero), all taken from our essays data. These projections largely align with an intuitive judgement of the gender-connotation of certain words, e.g. the words *dress* and *ladies* have a strong feminine association, whereas the words *army* and *team* have a strong male association.

To more formally validate the projection values from our WEM, we compare them to

the ratings of gender-connotations of 360 words by 540 psychology students in a study by Jenkins et al. (1958). We find a high correlation (0.52) between our projection values and the student ratings. Moreover, our method captures more variation in the gender-connotation of words than the ratings in Jenkins et al. (1958)'s study (see Appendix C.3 for details).

Figure 3: Projecting Words onto the Gender Vector



Note: The graph plots cosine similarities of the words on the y-axis with our gender dimension vector. Words with a cosine similarity between $[-0.1, 0.1]$ are neutral. Words with cosine similarity greater (less) than 0.1 (-0.1) are classified as feminine (masculine) words. Blue dots indicate the 15 most masculine words, green dots indicate 15 most feminine words, orange dots indicate 10 neutral words

Note that our WEM only assigns vector values to 1-grams, and therefore this procedure only assigns projection values to 1-grams. However, our Bag of Words representation also contains 2-grams which begin with either "not" or "no". For these phrases, we assign -2 times the projection value of the second word in the phrase. For example, "not marry" is assigned -2 times the projection value for "marry". This allows the effect of the presence of

'marry' to be removed and additionally allows an additional subtraction to reflect that the essay-writer mentions that they will not be married.

At this stage, we also drop neutral words from our Bag of Words matrix to focus on words with a meaningful gender association. We do this by standardizing the projection values for all words in our Bag of Words representation, and dropping those that have standardized projections less than one standard deviation away from zero. We also drop words which occur less than 200 times across our whole sample of essays. This is to avoid the index being driven by rarely used, or still misspelled, words. As we describe in Section 4.4, dropping these words has little impact on our headline results.

3.3.6 Step 6: Constructing an essay-level measure of gender conformity

We have now assigned a projection value to all 1-grams, and 2-grams which start with "no" or "not", across the essays. To construct an essay-level measure, we sum the counts of each word weighted by its projection value for each essay. We then regress this weighted sum on a quadratic of the number of words in each essay, and take the residual, to ensure longer essays do not mechanically have more extreme weighted sums. We standardize these residuals, and winsorize them at the 1st and 99th percentile to arrive at our index of gender conformity.

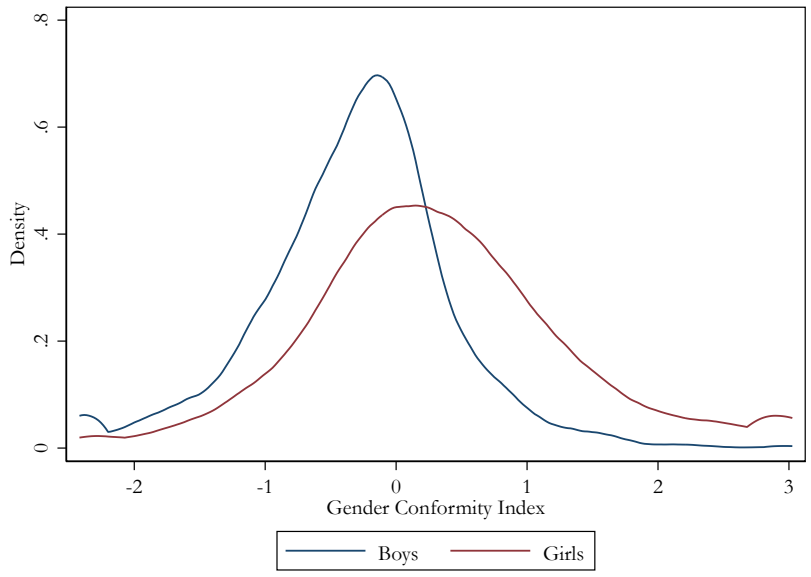
3.4 Describing our Index

For steps one to six, we use all essays available in the age 11 wave of the NCDS; that is, 10,511 essays written by both boys and girls. Figure 4 plots the distribution of our gender conformity index for the 5,419 boys and 5,091 girls who wrote essays (we do not know the gender of 1 writer). As one would expect, girls on average write more female-gendered essays; the mean of the index for girls is 0.33, which is statistically different to the mean of -0.33 for boys at the 1% significance level. Interestingly, girls' essays also display considerable dispersion in the gender conformity they convey. For our analysis sample of 3,497 girls, the index ranges from -2.42 to 3.03, with mean 0.33 and standard deviation 1.01.[15]

Our WEM model selects which words in the essays are most consequential for our index of gender conformity. Which words matter most will depend not only on how gendered they are according to our WEM, but also how frequently they appear in the essays. To investigate which words drive the variation in our gender conformity index for the full sample of 10,511 essays, we use a LASSO approach. We use our gender conformity index as the outcome

[15]There is less dispersion in the boys' essays. 82% of boys write essays with an index value within 1 standard deviation of 0. This is because most boys do not talk about topics, or use language, which is very gendered. Others write about a mix of feminine and masculine topics whose effects net out in our index.

Figure 4: Distribution of the Gender Conformity Index for boys and girls



Notes: This graph shows the distribution of the gender conformity index for boys and girls using the full sample of essays, that is 5,419 boys essays and 5,091 girls essays.

and counts of all essay words weighted by their projection values as the regressors. The ten most important words for predicting our index turn out to be: female, school, girl, married, husband, work, baby, one, new, and hair.[16] It is reassuring that most of these words relate to either traditionally masculine activities (e.g. work) or feminine ones (baby, married).

A natural question is whether a simpler approach of choosing and counting intuitively feminine and/or masculine words across essays would work just as well as our index. To test this, we construct an alternative index, which counts occurrences of a list of feminine words that we select independently of our WEM.[17] We find that this simpler index, based on an ad-hoc selection of words, has a correlation of 0.46 with our gender conformity index and explains 21% of our index's variation. Details of this comparison can be found in Appendix C.3. This suggests that the key advantage of our measurement procedure is that we account for *all* words with a possible gender-connotation in the essays; focusing on a subset omits plausibly relevant information on gender conformity. Moreover, as previously mentioned,

[16] We also try an alternative approach. We count the frequency of words which appear in essays that lie in the top and bottom 25% of the gender conformity index distribution. We then weight each word count by the weights from the word embedding model and calculate the difference in weighted frequency between the bottom 25% and the top 25% of the index distribution. The ten words for which weighted frequency is most different are: female, husband, girl, married, baby, hair, mother, school, team and girls.

[17] These words are: husband, house, boy, girl, little, baby, stay, cook, wash, clean, breakfast, knit, tidy, kitchen, housework, sew, hostess, wedding, washing, part, and let.

our procedure accurately reflects female connotations at the time the essays were written. Having a group of researchers choose a list of gendered words would likely be biased by their own (and current) experiences of gender norms.[18]

To further clarify what our gender conformity index captures, we revisit the essays in Section 2.2.

Essay 1: "*There was a pile of washing waiting to be done in the laundry basket waiting for me to wash. Oh how I wish I was young I thought to myself. Just as I had the water in the washer I heard my five month old baby crying in her pram outside. It was her bottle time I have to leave every thing to get her bottle ready. As soon a I had fed her my husband came home for his dinner. It had to be a ham (sand/samwhich) samwidgch today so I could get my washing done quicker. [...]*"

Essay 2: "*If I was 25 now, I would probably live in a small flat. I would go to work at 9 o'clock in the morning and finnish at 5 o'clock, I would make my own meals. After my tea I would go out in my car. When summer came round I would go away for a week say Paris or New York. My flat would have modern furniture in it. I would hold lots of parties and have beer, sausage on sticks, salad and Chinese food. [...]*"

Here, we highlight some words in these essays based on their gender association. Words in pink have a female association, while words in blue have a male association. Words in pink (blue) project positively (negatively) onto the gender dimension vector that we construct. Darker shades of pink and blue indicate stronger gender associations.

Consistent with a casual read of these essays, the first has a gender conformity index value of 1.46, while the second has a value of -0.11. The index's numerical value should be read as saying that the writer of the first essay, for example, uses words which are 1.13 standard deviations more feminine than the mean female essay (which equals 0.33). We therefore say the gender conformity of the writer of the first essay is 1.13 standard deviations stronger than the average girl in our sample. The sign of the index values indicate that the first essay is in line with traditionally female gender roles, while the second one is more associated with male gender roles. Overall, the first essay is also more strongly gendered than the second (as captured by the index's magnitude); this is evident from the highlighted words which are all pink in the first essay, but mixed in the second.

[18]Similarly, we refrain from using ChatGPT or other Large Language Models, as it is unclear what data it is trained on for this time period.

4 Results

In this section, we first show that there is a sizable association between lifetime earnings and gender conformity. We show this association is due to gender conformity's association with both labor supply and wages, and we study three channels through which our index predicts labor market outcomes: educational attainment, family formation, and occupational choice. Finally, we present results on potential determinants of gender conformity. To assess the sensitivity of our results to how we measure gender conformity, and to rule out competing interpretations of our findings, we include a wide range of robustness checks.

4.1 Lifetime Earnings and Gender Conformity

For earnings, labor supply and wages, we use OLS to estimate the following regressions:

$$y_i = \beta_0 + \beta_1 \text{GenderConformityIndex}_i + \beta_2 \text{Cog11}_i + \beta_3 \text{NonCog11}_i + \gamma' \mathbf{X_i} + u_i \tag{1}$$

where y_i is one of four outcomes: log lifetime earnings, log average hourly wage, log of total hours worked over the life cycle, or share of working life spent in employment. Our coefficient of interest is β_1. We report three specifications for our four main outcomes. The first is a univariate regression of the outcome on the gender conformity index. The second includes the indices of cognitive and non-cognitive skill measured at age 11, Cog11_i and NonCog11_i, which have been shown to be important for explaining lifetime earnings and human capital accumulation (Cunha et al., 2010). In our third (and main) specification, we additionally include our full vector of controls, $\mathbf{X}_i$, including: parents' age, parents' education, parents' income, parents' aspirations for the education of their child, number of siblings, sibling sex composition, birth order, family stability, and county fixed effects. u_i is the residual.

Column 1 in Table 3 reports the coefficient from a regression of log lifetime earnings on the index of gender conformity when not controlling for other variables. Having one standard deviation stronger gender conformity is associated with a 5.5% decrease in lifetime earnings. Part of this is due to the correlation of gender conformity with cognitive skills which are, on average, 0.1 standard deviations lower for those with one standard deviation stronger gender conformity. Controlling for cognitive and non-cognitive skills, the coefficient on our index falls to 2.9%. Upon including control variables and county fixed effects, we arrive at our headline result: a one standard deviation increase in gender conformity predicts a lifetime earnings decrease of 3.5% (significant at the 5% level), as shown in Column 3. The magnitude of this coefficient is 1.3 times the size of the one on non-cognitive skills (also

Table 3: Lifetime Earnings

	Log Lifetime Earnings	Log Lifetime Earnings	Log Lifetime Earnings
Gender Conformity Index	-0.055***	-0.029*	-0.035**
	(0.016)	(0.016)	(0.017)
Cognitive Skills		0.225***	0.201***
		(0.018)	(0.020)
Non-Cognitive Skills		0.026*	0.026
		(0.016)	(0.016)
Dad: Post Compulsory			0.053
			(0.046)
Dad: Some College			0.025
			(0.092)
Mum: Post Compulsory			-0.003
			(0.043)
Mum: Some College			0.165
			(0.118)
Log Parental Income			0.027
			(0.038)
Constant	12.223***	12.176***	11.671***
	(0.017)	(0.018)	(0.321)
Family Background			✓
County Fixed Effects			✓

Notes: Sample size is 3,497 girls. We report significance at the 1% (***), 5% (**), and 10% (*) levels with robust standard errors. The first column shows the univariate OLS regression of log lifetime earnings on our gender conformity index. Next, we include (standardized) cognitive and non-cognitive skills indices. The family background variables we include are parental income and education. We also control for (although do not display): number of and sex composition of siblings, birth order, mother's and father's age at birth, parental aspirations, and family stability. To see all coefficients displayed, see Table A.1 in Appendix A.1. We include 111 county fixed effects. All specifications have dummies for missing regressor observations.

measured in standard deviations), whose role in economic outcomes has been heavily studied in the literature (Cunha et al., 2010).

To understand whether this earnings result is due to wages or labor supply, we separately consider how gender conformity predicts lifetime average wages, total hours worked over the life cycle, and share of life cycle spent in employment. Table 4 shows results from regressions with these three outcomes as dependent variables. Panel A of Table 4 reports results for the log of average (hourly) wages over the life cycle. Conditional on the full set of controls, we

find that a one standard deviation increase in our gender conformity index predicts a 2.0% decrease in wages (significant at the 10% level), indicating that there are wage differences between those who conform more to prevailing gender norms than others. Panel B shows that a one standard deviation increase in our index predicts 1.6% fewer hours worked over the life cycle. This echoes Goussé et al. (2017)'s finding that stronger family attitudes increase time allocation towards home production in Britain. The hours result does not operate through the extensive margin: as indicated in Panel C, there is no association between our index and the share of life cycle spent in employment (once controls are added).[19]

[19]Results on the intensive margin, that is for log total hours worked over the life cycle, conditional on being employed, can be found in Appendix A.1, in Table A.4.

Table 4: Lifetime Wages, Hours Worked

	(1)	(2)	(3)
Panel A: Log Lifetime Average Hourly Wages			
Sample mean: 1.67			
Gender Conformity Index	-0.035***	-0.016	-0.020*
	(0.011)	(0.010)	(0.011)
Cognitive Skills		0.168***	0.139***
		(0.012)	(0.013)
Non-Cognitive Skills		0.012	0.013
		(0.010)	(0.011)
Panel B: Log Total Lifetime Hours Worked			
Sample mean: 10.6			
Gender Conformity Index	-0.022***	-0.015*	-0.016*
	(0.008)	(0.008)	(0.008)
Cognitive Skills		0.062***	0.057***
		(0.009)	(0.010)
Non-Cognitive Skills		0.011	0.010
		(0.008)	(0.008)
Panel C: Share of Lifetime Spent in Employment			
Sample mean: 0.81			
Gender Conformity Index	-0.007**	-0.005	-0.005
	(0.004)	(0.004)	(0.004)
Cognitive Skills		0.018***	0.015***
		(0.004)	(0.004)
Non-Cognitive Skills		0.008**	0.008**
		(0.004)	(0.004)
Parental Education & Income			✓
Family Background			✓
County Fixed Effects			✓

Notes: Sample size is 3,497 girls. We report significance at the 1% (***), 5% (**), and 10% (*) levels with robust standard errors. The first column shows the univariate OLS regression of each lifetime outcome on our gender conformity index. Column (2) includes indices for cognitive and non-cognitive skills. Column (3) includes our full set of family background and geographic controls, including (as displayed) parental income and education. We also control for (but do not display): number of and sex composition of siblings, birth order, mother's and father's age at birth, parental aspirations, and family stability. We include 111 county fixed effects. All specifications have dummies for missing regressor observations.

Table 5: Life Cycle Trends in Earnings

	Age 23	Age 33	Age 42	Age 50	Age 55
Panel A: Log Earnings in Survey Years					
Gender Conformity Index	-0.002	-0.056***	-0.047**	-0.018	-0.021
	(0.009)	(0.021)	(0.021)	(0.016)	(0.024)
Cognitive Skills	0.115***	0.223***	0.216***	0.196***	0.225***
	(0.011)	(0.023)	(0.024)	(0.021)	(0.027)
Non-Cognitive Skills	0.005	0.027	0.024	0.029*	0.007
	(0.009)	(0.019)	(0.019)	(0.017)	(0.023)
Panel B: Log Average Hourly Wages in Survey Years					
Gender Conformity Index	-0.008	-0.019	-0.039**	-0.003	-0.003
	(0.007)	(0.014)	(0.016)	(0.012)	(0.020)
Cognitive Skills	0.092***	0.174***	0.148***	0.171***	0.144***
	(0.009)	(0.016)	(0.019)	(0.015)	(0.022)
Non-Cognitive Skills	0.010	0.022*	0.015	0.031**	-0.005
	(0.008)	(0.013)	(0.014)	(0.012)	(0.018)
Panel C: Log Annual Hours Worked in Survey Years					
Gender Conformity Index	0.007	-0.019**	-0.001	-0.009	-0.008
	(0.004)	(0.008)	(0.007)	(0.007)	(0.008)
Cognitive Skills	0.013**	0.056***	0.031***	0.018**	0.041***
	(0.005)	(0.009)	(0.009)	(0.009)	(0.010)
Non-Cognitive Skills	-0.006	-0.005	0.005	0.003	0.002
	(0.005)	(0.008)	(0.007)	(0.007)	(0.008)
Panel D: Employment Status in Survey Years					
Gender Conformity Index	-0.007	-0.004	-0.003	-0.005	-0.003
	(0.008)	(0.009)	(0.007)	(0.007)	(0.009)
Cognitive Skills	0.069***	0.028***	0.010	0.013	0.005
	(0.009)	(0.010)	(0.008)	(0.008)	(0.011)
Non-Cognitive Skills	0.015*	0.011	0.004	0.016**	0.017*
	(0.008)	(0.009)	(0.007)	(0.007)	(0.009)
Parental Education & Income	✓	✓	✓	✓	✓
Family Background	✓	✓	✓	✓	✓
County Fixed Effects	✓	✓	✓	✓	✓

Notes: Sample size is 3,497 girls. We report significance at the 1% (***), 5% (**), and 10% (*) levels with robust standard errors. Each column reports coefficients from a regression with the full set of controls: cognitive and non-cognitive skills indices, parental income and education, number of and sex composition of siblings, birth order, mother's and father's age at birth, parental aspirations, family stability, and county fixed effects. We include dummies for missing regressor observations.

While our main results are for outcomes aggregated over the life cycle, Table 5 shows disaggregated results by age. We find that our index predicts particularly large declines in earnings at age 33 and 42, whereas at younger and older ages the predicted declines are much smaller. These lower earnings are driven by fewer hours worked, especially at 33, and lower wages at 33 and 42. Age 33 and 42 are exactly the ages when the burdens of childcare are the largest. Both Adda et al. (2017) and Kleven et al. (2019) show that childbearing is associated with reductions in employment and earnings, and that the earnings declines are persistent, potentially due to the loss of human capital from work interruptions.

4.2 Channels

We have established that lower gender conformity is predictive of higher lifetime earnings, through both higher wages and, to a lesser extent, through more labor supplied over the life cycle. We now explore *why* gender conformity could be predictive of earnings by exploring some channels through which gender conformity could influence wages and labor supply. We examine how three key decisions made by cohort members – educational attainment, family formation, and occupational choice – are predicted by gender conformity, and how these, in turn, mediate the results for lifetime earnings, average hourly wages and total hours worked.

Educational Attainment One explanation for the lifetime earnings and average wage results is that girls with stronger gender conformity tend to attain less education, and therefore command lower wages than girls with less gender conformity. Panel A of Table 6 supports this explanation. We report regressions of educational attainment on the gender conformity index, controlling for our full set of covariates from Column 3 of Table 3 (including cognitive and non-cognitive skills, measured at the age of 11). For the years of education regression, the coefficient on gender conformity is -0.077, which indicates that a one standard deviation increase in gender conformity is correlated with a reduction of approximately one month spent in education. These declines occur across the educational distribution: a 1 standard deviation increase in the index predicts a 1.6 percentage point increase in the high school dropout rate (relative to a mean of 21 percentage points) and a 1.1 percentage point decline in college attendance (relative to a mean of 15 percentage points).[20] All three coefficients are statistically significant at the 5% level.

Family Formation We explore two features of family formation decisions which could be predicted by gender conformity, and which subsequently could explain our earnings results.

[20]We define high school dropouts as those who left school with qualifications lower than O-levels, and college attendance as those who attained a qualification above A-levels.

The first two columns of Panel B in Table 6 show regressions of early adulthood family formation outcomes – that is, whether the cohort member is married and has a child at 23. The latter two columns consider whether an individual ever got married, and the number of children they had by age 50 (to measure completed fertility). We can see that those who have stronger gender conformity tend to marry and have children early. A one standard deviation increase in gender conformity predicts a significant 2.1 percentage point increase in marriage (with a sample average of 55 percent points) and 1.4 percentage point increase in the proportion having had at least one child at age 23 (with a sample average 29 percent points). However, we find no statistically significant results on whether individuals ever get married or ever have children. Thus gender conformity largely predicts changes in the *timing* of family formation, which a large literature has found to come at the cost of lost career earnings (Adda et al., 2017; Kleven et al., 2019).

Occupational Choice Finally, we consider whether gender conformity can predict one's occupational choice, which could have implications for both labor supply and wages. Girls with stronger gender conformity are less likely to enter professional occupations, which includes engineers or legal professionals.[21] A one standard deviation increase in gender conformity predicts, on average, a 1.6 percentage point decline probability of being in such a profession. Instead, girls with stronger gender conformity are more likely to enter manual occupations, entailing jobs such as hairdresser and sales assistants[22]. Since professional occupations were much better paid than manual ones in Britain,[23] the lower earnings predicted by gender conformity could also in part be explained by differences in occupation choice.

[21]Specifically, it includes natural scientists, engineers, technologists, health professionals, teaching professionals, legal professionals, business and financial professionals, architects, town planners, and librarians.

[22]Specifically, it includes catering, travel attendants, childcare-related occupations, hairdressers, beauticians, domestic staff, sale and customer service occupations, buyers, brokers, and machinery operatives.

[23]Median female annual earnings of professionals in 2000 (when our cohort was age 42) were £22,546 versus £7,233 for manual jobs. This is computed using ONS earnings data for 2 digit SOC codes.

Table 6: Predicting Education, Family Formation and Occupation

	(1)	(2)	(3)	(4)
Panel A: Education				
Outcome:	**Years of Education**	**HS Dropout**	**Attend Uni**	
Model:	OLS	Logit	Logit	
Gender Conformity Index	-0.077***	0.016**	-0.011**	
	(0.026)	(0.006)	(0.005)	
Cognitive Skills	0.753***	-0.170***	0.116***	
	(0.031)	(0.008)	(0.006)	
Non-Cognitive Skills	0.081***	-0.016***	0.009*	
	(0.025)	(0.006)	(0.006)	
Sample Mean:	18.2 years	21.5%	14.8%	
Panel B: Family Formation				
Outcome:	**Marry at 23**	**Children at 23**	**Ever Marry**	**Total # of Children**
Model:	Logit	Logit	Logit	Logit
Gender Conformity Index	0.021**	0.014*	0.007	0.009
	(0.009)	(0.008)	(0.006)	(0.007)
Cognitive Skills	-0.028***	-0.084***	-0.003	-0.030***
	(0.010)	(0.009)	(0.007)	(0.009)
Non-Cognitive Skills	0.006	-0.019**	0.002	0.003
	(0.009)	(0.007)	(0.006)	(0.007)
Sample Mean:	54.8%	29.3%	87.0%	82.9%
Panel C: Occupation				
Outcome:	**Professional**	**Administrative**	**Manual**	**Elementary**
Model:	Multi-Logit	Multi-Logit	Multi-Logit	Multi-Logit
Gender Conformity Index	-0.016**	-0.009	0.016*	0.010
	(0.007)	(0.009)	(0.008)	(0.007)
Cognitive Skills	0.081***	0.067***	-0.078***	-0.070***
	(0.008)	(0.010)	(0.010)	(0.009)
Non-Cognitive Skills	0.013	-0.002	-0.006	-0.006
	(0.007)	(0.009)	(0.008)	(0.006)
Sample Mean:	20.3%	36.3%	28.0%	15.4%
Parental Education & Income	✓	✓	✓	✓
Family Background	✓	✓	✓	✓
Geographic Fixed Effects	✓	✓	✓	✓

Notes: Sample size is 3,497 girls. We report significance at the 1% (***), 5% (**), and 10% (*) levels with robust standard errors. For the OLS regressions, we report raw coefficients on the gender conformity index. For the binary logit regressions, we report average marginal effects. Panel C is estimated using a multinomial logit and reports predicted marginal effects for each category. Every column reports results from a regression with the full set of controls: cognitive and non-cognitive skills, parental income and education, number of and sex composition of siblings, birth order, mother's and father's age at birth, parental aspirations, family stability, and geographic fixed effects. For Panels A and B, we include county fixed effects. For Panel C, we use region fixed effects to ensure identification since there are counties for which we do not observe variation in occupation. We include dummies for missing regressor observations.

Table 6 suggests that gender conformity is predictive of lower educational attainment, earlier family formation and selection into lower-paid occupations. Table 7 quantifies the extent to which these three channels mediate our lifetime earnings result. We use a Gelbach decomposition to show the estimated contribution of each of these three channels to the coefficient on gender conformity, in a way that is robust to the sequence in which they are included. Details on the decomposition can be found in Appendix D and Gelbach (2016).

The first column ("Baseline") in Table 7 contains the coefficients on gender conformity from our main specifications reported in the third columns of Tables 3 and 4. The second column reports the coefficients from regressions which additionally include our mediators: education, family formation at age 23, and occupation. The subsequent columns report the shares of the baseline coefficients that are explained by education, family formation, and occupation, respectively. The final column shows the total shares of the baseline coefficients that are explained by these three mediators.[24]

Table 7: Effects Mediated through Education, Family and Occupational Choice

	Coefficient		Education	Family Formation	Occupation	Total
	Baseline	With Mediators				
Log Lifetime Earnings	-0.035**	-0.016	18.8%	4.6%	27.4%	50.8%
Log Avg. Hourly Wage	-0.020*	-0.007	31.0%	-1.3%	30.9%	60.6%
Log Total Hours Worked	-0.016*	-0.010	4.4%	14.0%	21.6%	40.0%
Employment	-0.005	-0.003	9.3%	12.3%	21.0%	42.5%

Notes: Sample size is 3,497 girls. We report significance at the 1% (***), 5% (**), and 10% (*) levels with robust standard errors. The first column reports our baseline coefficients, as presented in Tables 3 and 4. The second column additionally controls for education (years of school), family formation (married at 23 and children at 23), and occupation (indicators for 9 occupation groups). The "Education", "Family Formation", and "Occupation" columns report shares explained by each mediating group of variables. In each regression of the decomposition, we control for: cognitive and non-cognitive skills, parental income and education, number of and sex composition of siblings, birth order, mother's and father's age at birth, parental aspirations, family stability, and county fixed effects. Since missing dummies explain a small fraction of the association between gender conformity and outcomes, the total share explained in the last column does not exactly equate the percentage change between the "Base" and "With Mediator" column.

Comparing estimates from the models with and without mediators reveals that approximately half of the association between gender conformity and lifetime earnings is explained by education, family formation, and occupation. Furthermore, our mediators explain 61%,

[24] In theory, the total shares are equivalent to the percentage change in coefficients between the baseline model and the model with mediators, though there is a slight discrepancy in our case, due to the additional missing value dummies which we include in the saturated model.

40%, and 43% of the association with wages, hours, and employment, respectively. Education and occupation are the main drivers of the earnings and the wage results; for example, educational attainment and occupation explain 19% and 27% of the earnings result. On the other hand, when it comes to the reduced working hours of those with stronger gender conformity, earlier family formation and occupational choice are the most important drivers. To summarize: women with stronger gender conformity attain less education, which reduces their wages, and have children earlier, which reduces their working hours. Occupations chosen by those with stronger gender conformity predict lower wages and fewer hours worked.

4.3 Predictors of Gender Conformity

Gender conformity exhibits significant associations with lifetime outcomes. A natural next question is: what determines how much a girl conforms with gender norms? Our setting is a particularly interesting one to ask this question, because we can examine the role of micro-determinants of gender conformity, among a sample of girls who experience a relatively homogeneous upbringing (they are all white and born in Britain in 1958).

To study this question, we regress the gender conformity index on a large set of potential determinants, including child skills, parental education, family demographics, role model variables, and region-level characteristics. Figure 5 reports on the left-hand side the coefficients of a univariate regression of our index on each determinant, and on the right-hand side reports the coefficients from a regression in which we include all determinants. In the following, we focus our discussion on the results from the saturated model.

Children with higher age-11 cognitive and non-cognitive skills tend to exhibit lower gender conformity. Once we control for the full set of covariates, a one standard deviation improvement in cognitive skills is associated with 0.08 standard deviations lower gender conformity; for non-cognitive skills it is associated with 0.04 standard deviations lower gender conformity. Table A.2 in Appendix A.1 displays these coefficients. Parental income does not seem to matter for gender conformity; the coefficient is indistinguishable from zero.[25]

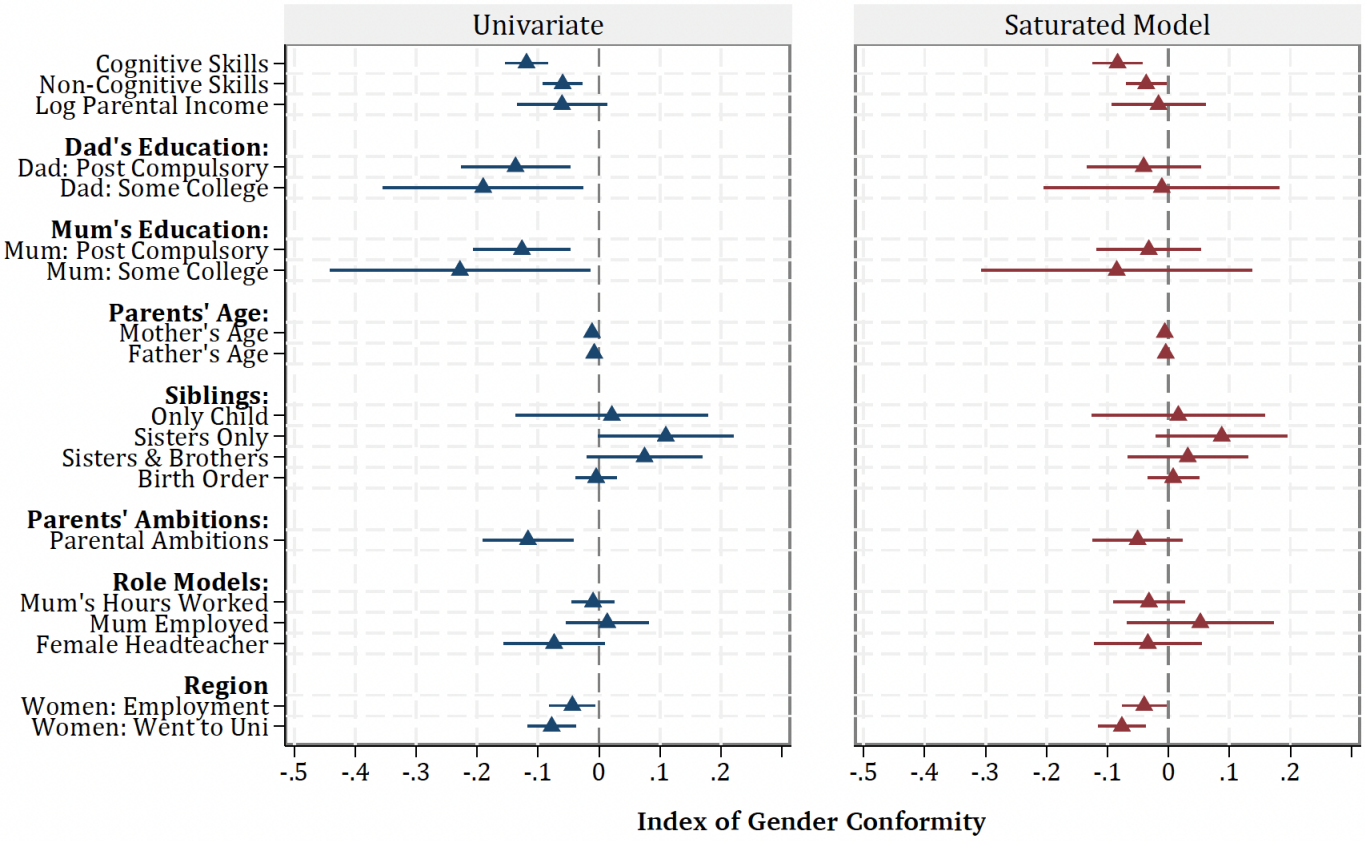
Parental education is statistically significant in the univariate regressions, but not in the saturated regressions. The point estimates suggest that having more educated parents is associated with being less gender conforming. This is similar to recent evidence that parents play an important role in transmitting gender attitudes to their children (Dhar et al., 2019).

Previous literature has found that the number and composition of siblings can poten-

[25]This result is robust to using other specifications; for example, being at the top or bottom quartile of parental income, compared with being in the middle of the distribution, does not predict any difference in gender conformity.

tially influence gender conformity by affecting gender roles within the household (Brenøe, 2022). We find evidence (albeit noisy) that, relative to having only brothers (who are the reference group in the regression), girls who have only sisters tend to display stronger gender conformity. Those with no siblings tend to gender conform similarly to those with only brothers. In short, the presence of sisters in a household seems to predict gender conformity.

Figure 5: Determinants of Gender Conformity Index



Notes: Sample is 3,497 girls. Triangles represent point estimates, bars represent 95% confidence intervals, computed using robust standard errors. Saturated model coefficients are from a regression of the gender conformity index on all the variables shown in the figure, plus indicators for missing observations.

The age 11 NCDS survey asks whether parents hope their children will undertake further education or training upon leaving school. We find a negative association between this measure of parental ambition and gender conformity, albeit imprecisely estimated. Parents who hope their daughter continues education beyond the compulsory school-leaving age tend to have girls with 0.1 standard deviations lower gender conformity. This may indicate some form of inter-generational transmission of preferences or beliefs.

Role models have been shown to matter for individuals' attitudes towards gender roles

(Fernández et al., 2004). We find that having an employed mother, and the number of hours a mother works, are not strongly associated with gender conformity. Similarly, having a female headteacher does not predict how gender-conforming a girl is. Since the NCDS data does not contain detailed information on a cohort member's school, a potential role model effect might be counteracted here by differences in the types of schools female head teachers lead (e.g. they might be more likely to head a girls' only school).

Finally, we consider geographic differences. We find that growing up in an area with a higher fraction of women who work is associated with lower gender conformity of girls. Similarly, in areas where a larger proportion of adult women attended university relative to men, girls tend to display lower gender conformity. Aside from cognitive and non-cognitive skills, these geographic differences in women's educational attainment and employment are the only statistically significant predictors of girls' gender conformity we find when we jointly consider all potential determinants. These findings are in line with many papers studying gender attitudes more broadly, which assume that women coming from places with traditional gender norms also adhere more to such roles in their decisions (Fernandez, 2007).

Column 1 of Table A.3 in Appendix A.1 reports Shapley values, which give the share of the R^2 in the saturated model attributable to a given variable. The R^2 from this regression is 0.035. Our Shapley values affirm that cognitive and non-cognitive skills, as well as the fractions of women working and who went to university in a given region, are the most important determinants of gender conformity amongst the ones we examine in our data. This finding is further reinforced when we use a simple regression random forest to predict our gender conformity index – the second column in Table A.3 displays variable importance from the random forest, showing that exactly these four variables have the most predictive content for gender conformity. While our ability to predict gender conformity is improved if we use a simple random forest, as opposed to a linear OLS regression, we note that much of the variation in the index is still unaccounted for.[26]

In summary, whilst one might have thought that gender conformity is shaped primarily by families, we do not find evidence for this – apart from individuals' skills, the role of parents and the family appears to be minor, and instead the wider environment girls grow up in matters. This is an interesting avenue for future work.

[26]The RMSE is 0.78 when using a simple random forest and is 1.00 for the saturated model we present in Figure 5.

4.4 Robustness Checks

In this sub-section, we conduct a number of robustness checks related to: concerns of omitted variables, potential misspecification of our main regression specification, different ways of constructing the gender conformity index, as well as measurement error in the index. We also report how gender conformity predicts labour-market outcomes for our sample of boys.

Other Latent Information Contained in the Essays A natural concern is that our index potentially captures unobservables that influence both earnings and gender conformity. For example, educational (or other) aspirations may also be correlated with gender conformity. To consider this possibility, we construct six indices to capture further content of the essays using the same natural language processing methods as before. Following Kozlowski et al. (2019)'s study of social class, the dimensions we use to construct the six indices are: (1) education (educated-uneducated), (2) cultivation (cultured-uncultured), (3) affluence (rich-poor), (4) status (prestigious-undistinguished), (5) morality (good-evil), and (6) employment (employer-employee). For each of these dimensions, we construct an index analogously to the gender conformity index.[27] We project each word in each essay on each of the above six dimensions, and take the aggregated projection values (regression-adjusted for essay length) as the index for each dimension. The results from controlling for these six additional indices are reported in Panel A of Table 8. Our main results, based on the regression specification in equation (1), are robust to the inclusion of the other dimensions. In particular, one standard deviation more gender conformity still predicts a 3.0% decrease in lifetime earnings, a 1.6% decrease in average wages and a 1.5% decrease in total hours worked over the life cycle, even after controlling for the other indices (although the estimates are noisier, which is to be expected as the indices are correlated). None of the other dimensions significantly predict lifetime outcomes except for the education index predicting wages and the employment index predicting share of life cycle spent in employment. This affirms that we indeed capture gender conformity rather than other latent content in the essays, and that gender is the key dimension in which our essays are informative.

[27] For example, for the education index we first exactly follow steps 1 to 3 as given in Section 3.3. We then construct an education dimension vector, using the following pairs of antonyms: educated-uneducated, learned-unlearned, knowledgeable-ignorant, trained-untrained, taught-untaught, literate-illiterate, schooled-unschooled, tutored-untutored and lettered-unlettered. We then assign projection values for education to words in the essays, by projecting them onto the education dimension vector. We aggregate weighted word counts (as in step 5 in Section 3.3), to arrive at an education index. Details on the other dimensions are in Appendix C.4.

Oster Bounds There is still the broader concern that our gender conformity index is predictive of lifetime outcomes because we omit unobservables that may be correlated with gender conformity. One way in which we address this is by reporting Oster bounds for our four main coefficients in Panel B (Altonji et al., 2005; Oster, 2019). The approach is based on the idea that the extent of selection on observables can be used to infer the extent of selection on unobservables. In particular, the extent of variation in unobservables, as well as their correlation with the outcome of interest, can be inferred from a saturated model that includes all observables. These bounds are constructed using Oster (2019)'s recommended parameter values.[28] Reassuringly, our bounds are not large and never contain zero, implying that our main results would not be overturned even if unobservables in our study were as important as observables. The second row in Panel B reports "δ", the degree of selection on unobservables relative to observables which would be necessary to explain away each result. In all cases, selection on unobservables would need to be at least thrice, if not up to seven times, as important as selection on observables to find that gender conformity predicts a zero change in each lifetime outcome.

Complementarities between Predictors Another concern may be misspecification. For example, our results may be driven by complementarities between the control variables, which are unaccounted for in our linear specifications. To investigate this possibility, we include quadratics and interactions of all covariates in our model and re-estimate it using a post-LASSO OLS regression, which eliminates terms with low predictive power. Panel C in Table 8 shows that accounting for non-linearity between predictors leads to slightly larger estimated declines in earnings, wages, and labor supply predicted by gender conformity.

Additional Controls for Skills, School Quality or Aspirations Our index could be proxying for other important predictors of lifetime outcomes. In Table A.6, we explore whether our results are robust to the inclusion of: (i) non-linear functions of cognitive and non-cognitive skills, (ii) additional controls for cognitive skills (i.e. essay length and spelling mistakes), (iii) quality of the school the girl attends at the time of writing the essay, and (iv) the occupation the girl declares to aspire to in the survey questionnaire. In each case, our results are very similar to the baseline estimates in Tables 3 and 4.

[28] The lower bounds are the coefficients from our main specifications. To calculate the upper bounds, we assume (1) equal selection on observables and unobservables and (2) the R^2 from a hypothetical regression of the outcome on treatment and both observed and unobserved controls would be 1.3 times the R^2 from our main specification.

Table 8: Alternative Explanations

	Lifetime Earnings	Lifetime Average Wages	Employed	Log Total Hours Worked
Panel A: Other Text Factors				
Gender Conformity Index	-0.030*	-0.016	-0.005	-0.015*
	(0.017)	(0.011)	(0.004)	(0.009)
Education Index	0.019	0.026**	0.000	-0.003
	(0.019)	(0.012)	(0.004)	(0.010)
Cultivation Index	0.012	0.008	0.001	0.003
	(0.022)	(0.014)	(0.005)	(0.011)
Affluence Index	-0.003	0.001	0.000	-0.001
	(0.016)	(0.011)	(0.004)	(0.008)
Status Index	-0.019	-0.013	-0.006	-0.006
	(0.017)	(0.011)	(0.004)	(0.008)
Morality Index	-0.008	-0.010	-0.003	-0.005
	(0.020)	(0.013)	(0.005)	(0.010)
Employment Index	0.007	0.012	-0.007*	-0.008
	(0.018)	(0.012)	(0.004)	(0.009)
Panel B: Oster Bounds				
Bounds on Gender Conformity Index	[-0.035, -0.028]	[-0.020,-0.014]	[-0.006,-0.005]	[-0.016, -0.014]
Delta (δ)	4.38	3.45	7.04	6.12
Panel C: Complementarities - Post-LASSO Model				
Gender Conformity Index	-0.035**	-0.021**	-0.008**	-0.023***
	(0.016)	(0.011)	(0.004)	(0.008)
Cognitive & Non-Cognitive Skills	✓	✓	✓	✓
Parental Education & Income	✓	✓	✓	✓
Family Background	✓	✓	✓	✓
County Fixed Effects	✓	✓	✓	✓

Notes: Sample size is 3,497 girls. We report significance at the 1% (***), 5% (**), and 10% (*) levels with robust standard errors. Every column and panel report prediction results from regressions with the full set of controls: cognitive and non-cognitive skills, parental income and education, number of and sex composition of siblings, birth order, mother's and father's age at birth, parental aspirations, and family stability. We also include 111 county fixed effects. Panel A adds indices for six latent factors, also derived from our essays: education, cultivation, affluence, status, morality, and employment. Panel B reports Oster bounds for our main coefficient. The first row shows bounds on the coefficient on the gender attitude index, with R_{max} equal to 1.3 times the R^2 from Column 3 of Table 3. The second row reports the level of selection on unobservables relative to selection on observables (δ) that there would have to be for the coefficient to be equal to zero (Altonji et al., 2005; Oster, 2019). Panel C performs an OLS regression on LASSO-selected covariates – we only report the index, which is always selected by LASSO (unlike non-cognitive skills). We include dummies for missing regressor observations.

Measurement Error in Gender Conformity Recall that our procedure for estimating gender conformity relies on a gender dimension vector, which is an average of vector values of ten gender-related word pairs. This averaging procedure reduces measurement error relative to using one particular pair of words (Kozlowski et al., 2019). To assess the role of

measurement error, we construct ten different gender conformity indices, each of which is based on only one of our ten pairs of gendered words (as opposed to the average of all ten). We then use these ten gender conformity indices as instruments for our original index. This is similar to Agostinelli and Wiswall (2016a)'s approach of using one measure of cognitive skill as an instrument for another. Appendix Table A.5 shows that when instrumenting, the coefficient on gender conformity in the lifetime earnings regression rises slightly, while coefficients for the wage and hours worked regressions are slightly attenuated. This affirms that measurement error arising from our gender dimension based on 10 word pairs is unlikely to be of first-order importance.

Varying Index Construction In Appendix A.2, we present several robustness checks related to how we constructed our gender conformity index. First, we explore whether our results still hold if we use pre-trained word-embedding models trained on modern text, rather than our own model trained on text written between 1958 and 1978. To do so, we re-estimate our index based on Google's Word2Vec and Stanford's GloVe word embedding models, which are trained on extensive corpora of predominantly modern, digitised texts. Table A.7 shows that our qualitative results do not change. However, when we use models trained on modern text, the results are attenuated. This is consistent with findings in Kozlowski et al. (2019), who show that gender connotations in English have evolved throughout the twentieth century. For example, "nurse" was a female-gendered word in our 1958 to 1978 sample, but does not have a strong female association in Google's Word2Vec WEM and the GloVe WEM. The second check is regarding the construction of the index. We investigate robustness to our decisions to drop words which do not appear at least 200 times across essays and to drop words less than one standard deviation away from gender neutrality. Table A.8 confirms that our results are remarkably robust to these choices.

Results for Boys Appendix Table A.9 shows results when repeating our main analysis for the 3,513 boys for whom we observe essays and outcomes. We use the same gender conformity index as for girls; that is, essays with a larger number of female-associated words have a higher index value than essays with more male-associated words. Thus, if using more masculine language is associated with higher earnings for men, one should expect a negative coefficient. Controlling for skills, our index has a near-zero and insignificant association with each of lifetime earnings, average wages, share of life cycle spent in employment, and total hours worked for men. This is despite the fact that our gender conformity index not only captures female-associations of words, but also male-associations. Thus, conditional on skills, we do not find evidence that boys who use particularly masculine words do better in the

labor market, nor do we find evidence that boys who use particularly feminine language do worse, as suggested by Banan et al. (2023). We view this as evidence that gender conformity may be an important driver of labor market decisions for women, but not for men.

4.5 Summary of Results

Stronger gender conformity predicts that a girl will earn less over her lifetime. This earnings result is mostly driven by gender conformity predicting lower wages as well as fewer hours spent working over the life cycle. We argue that our results are not attributable to our index of gender conformity confounding other latent information in the essays, proxying for skills, or capturing complementarities between other predictors. Conditional on an extensive list of controls, stronger gender conformity is also associated with lower educational attainment, earlier childbearing and marriage, and selection into lower paid occupations. These intermediate decisions explain just over half of the earnings result we report.

5 A Model of Gender Conformity

Our empirical results speak to the role of gender conformity in predicting women's labor market outcomes. However, we have so far been agnostic about why gender conformity may influence these outcomes, and thus how to interpret our gender conformity index. In this section, we use a simple model to present two potential interpretations of why gender conformity may affect economic choices, and show that both are consistent with our results.

5.1 Interpreting our Index

One could interpret our index of gender conformity as measuring a girl's preference to engage in activities and behaviors which are closer to the roles assigned to women in society at the time. For example, there might be heterogeneity in the utility from home production (such as cooking and housework), which girls would write about in their essays. This preference would also play a role in whether they choose to go to university or the hours they work as adults. A second interpretation of our index is that it measures the different (perceived) constraints faced by young girls. For instance, some girls may live in areas with more conservative gender norms, and thus face higher costs of going to university because of discrimination or lack of support from their parents. Such constraints could prompt these girls to imagine themselves in more traditional gender roles in the future, despite this not being their preference.

We present a simple model demonstrating that both interpretations are consistent with the empirical patterns we observe. We cannot separately identify the extent to which the

associations we observe derive from preferences versus constraints. However, under either interpretation, our empirical findings add to our understanding of women's human capital investments and labor supply decisions. If we capture preferences, then we make observable an important driver of female labor supply decisions. For example, Adda et al. (2017), formulate a dynamic model of female labor supply and fertility choices. To match variation in observed choices, they include unobserved heterogeneity in preferences for children. Such differences in preferences would be directly captured by our gender conformity index. If instead the observed patterns reflect constraints, then we show that even among a relatively homogeneous sample, some girls face (or perceive) very different gender-based constraints by age 11. As Section 4.3 finds, these likely arise from environments outside the home.

5.2 Model Set-up

Our model nests either the preferences or the constraints interpretation of gender conformity. In the model, women live for two periods - youth and adulthood. In the first period (youth), they choose how to devote their time to leisure, education, and work, as well as how much to consume. One should think of this as capturing the work and education choices made between ages 16 and 23 in our survey. In the second period (adulthood), the woman chooses how much time to devote to leisure, home-production and work (for a wage that is increasing in her first period education decision), and how much to consume. In both periods, an individual's choices are subject to time and budget constraints. For tractability, we abstract away from occupational and family formation choices.

Our model has two key parameters which allow for the possibility of heterogeneous constraints and preferences among women. The first is γ, which captures a woman's preference for home production. Preferences for home production, which includes time spent with children and doing housework, have received attention in several recent papers on the gender earnings gap (Gayle and Golan, 2012; Gayle and Shephard, 2019; Boerma and Karabarbounis, 2021). Our index of gender conformity plausibly proxies for preferences over home production, especially since much of the index's variation comes from words relating to domestic chores and children (see Section 3.4). The second parameter is ξ which captures the (time) costs of attaining education; girls may be observed to conform with prevailing gender norms because the costs (e.g. due to unsupportive parents or discrimination among peers) of not doing so are high (Kozlowski and Power, 2022).[29]

[29] An alternative interpretation of constraints faced by women would be that they anticipate labor market discrimination in period 2. We could model ξ as the wage-penalty, for a given level of education, that women face – this would have isomorphic implications to the ones we find with ξ as a period 1 time-cost.

We show below that, in this simple set-up, a stronger preference for home-production or higher time costs of education will both imply: (i) less education attained and (ii) less market work in adulthood relative to time spent in home-production. These implications are consistent with what we find in the data.

5.2.1 Period 2: Adulthood

In period 2 (adulthood), women choose consumption, as well as time allocated to work, home-production and leisure. They maximize

$$\begin{aligned} V_2(ed;\gamma,\xi) &= \max_{c_2,l_2,hp_2} u(c_2,l_2,hp_2;\gamma,\xi) \\ &= \max_{c_2,l_2,hp_2} \frac{1}{\tau}\ln[(1-\gamma)c_2^{\tau}+\gamma hp_2^{\tau}]+\delta\ln(l_2) \end{aligned}$$

subject to:

$$\begin{aligned} T &= hp_2+l_2+h_2 \\ c_2 &= h_2\cdot w_2(ed) \end{aligned}$$

where ed is the level of education attained in youth, c_2 is consumption of market goods, l_2 is time allocated to leisure, h_2 is time spent working and hp_2 is time spent in home-production. In adulthood, the wage function $w_2(ed)$ is increasing and concave in educational attainment ed (i.e., $w_2'(ed)>0$, $w_2''(ed)\leq 0$), which is a common assumption (Ben-Porath, 1967).

Preferences for home production, as proxied by our gender conformity index, enter through $\gamma\in(0,1)$. Constraints on educational attainment do not directly enter the value function in period 2. Instead, they operate through $w_2(ed)$, since ξ affects educational choice in period 1. There are two additional parameters, which we assume are fixed across women. The first is $\tau\in(-\infty,1)$, which governs the elasticity of substitution between consumption of market and home-produced goods. The elasticity is given by $1/(1-\tau)$. The second is δ, the utility of leisure weight.

5.2.2 Period 1: Youth

In period 1, girls choose how much education to attain and how much to work. They maximize

$$\begin{aligned} V_1(\gamma, \xi) &= \max_{ed, c_1, l_1} u(c_1, l_1) + \beta V_2(ed; \gamma, \xi) \\ &= \max_{ed, c_1, l_1} \ln(c_1) + \delta \ln(l_1) + \beta V_2(ed; \gamma, \xi) \end{aligned}$$

subject to

$$\begin{aligned} T &= l_1 + \xi ed + h_1 \\ c_1 &= h_1 \cdot w_1 \end{aligned}$$

where ed is the level of education attained, c_1 is consumption of market goods, l_1 is time allocated to leisure, h_1 is time spent supplying labor in the market. In youth, the wage is fixed at w_1. We assume education has a marginal cost which is potentially heterogeneous across women and is given by ξ; this can also be interpreted as a cost of deviating from gender norms regarding education (akin to the Akerlof and Kranton (2000) model of identity). Education has no associated monetary cost.

5.3 Model Implications

We focus on preferences versus constraints as alternative interpretations of our gender conformity index by studying how our two key parameters, γ and ξ, affect education and labor supply choices. We derive the following optimality condition from the youth problem:

$$\beta \frac{\partial V_2(ed; \gamma, \xi)}{\partial ed} = \frac{\partial u(c_1, l_1)}{\partial c_1} \xi w_1 \tag{2}$$

which says that the marginal benefit of education (through higher wages in adulthood) equals the marginal cost of lost wage income during youth.

Given optimal choices in adulthood, we can rewrite equation (2) as[30]:

$$w_2(ed)^{\frac{2\tau-1}{1-\tau}} \left[\beta \frac{1}{\xi} h_1 w_2'(ed) - w_2(ed) \right] = \left(\frac{\gamma}{1-\gamma} \right)^{\frac{1}{1-\tau}} \tag{3}$$

The LHS is decreasing in education (as each of its elements is decreasing in education), under our assumption that $w_2(\cdot)$ is increasing and concave in education. The RHS is increasing in γ since it is between 0 and 1 and τ is less than 1. Thus, a girl with stronger preferences for home production (higher γ) will spend less time in education. These implications are

[30] See Appendix E for details.

identical if we instead model our index of gender conformity as a constraint. If educational attainment is more costly (higher ξ), this too will reduce educational attainment.

Combining first-order conditions from the problem in adulthood yields (the full solution is given in Appendix E):

$$\frac{h_2}{hp_2} = w_2(ed)^{\frac{\tau}{1-\tau}} \left(\frac{1-\gamma}{\gamma}\right)^{\frac{1}{1-\tau}}. \tag{4}$$

Suppose there is heterogeneity only in preferences for home production. Since the RHS is decreasing in γ, women with stronger preferences for home production work fewer hours relative to hours spent in home production. This effect is reinforced by the fact that these women also attained less education in their youth, and therefore receive a lower market wage, $w_2(ed)$. A larger τ, which is associated with a lower elasticity of substitution, accentuates both these channels; women with strong preferences for home production (and therefore with less education), work even less relative to time spent in home production if they are also less willing to substitute home-produced goods with market consumption.

Now consider heterogeneity in constraints across women. A higher ξ, that is a woman for whom time spent in education is more costly, implies more hours spent in home production relative to market work. Indeed, the impact of ξ is observationally equivalent to that of γ. The only difference is that ξ's effect in adulthood only operates through the wage channel – higher ξ implies a woman is less productive in adulthood due to lower education, but does not directly influence her time use. In contrast, higher γ affects educational decisions and thus productivity, as well as labor supply choices conditional on an individual's productivity.

Our simple model matches several key patterns in the data. The first is that girls with more gender conformity – that is, girls with a stronger preference for home production or higher costs of education – attain less education. This is what we show in Table 6. Women with higher gender conformity also tend to work fewer hours over their life cycle as seen in Table 4. Furthermore, women with stronger gender conformity are found to spend more time in home production activities, which is consistent with our results in Table 6 showing that our index predicts marriage and child-bearing at a younger age (which arguably represents spending a greater share of adult life in home production).

The structure of our model means that our insights are isomorphic to whether we interpret our index of gender conformity to be a measure of γ or ξ. We leave it for future research to disentangle the role of these two potential mechanisms in generating our results.

6 Conclusion

In this paper, we use natural language processing and novel text data from over ten thousand essays written by 11-year-old children, to construct an index of gender conformity for a representative sample of girls born in Britain in 1958. We link this index to data on childhood circumstances and outcomes over the life cycle, which allows us to study the determinants as well as the consequences of gender conformity.

Whilst many papers have studied gender attitudes more broadly, a much smaller literature studies the extent to which girls internalize gender norms and how this relates to their labour-market outcomes. Our approach has several important strengths. First, we observe gender conformity in childhood, meaning that concerns of reverse causality or justification bias are minimal. Second, by training our own word embedding model on over 2 million books written between 1958 and 1978, we can measure contemporaneous gender associations of words and phrases, and thus estimate girls' conformity to gender norms prevalent at the time the essays were written. Third, we can link the essays to rich data in childhood as well as rich data throughout adulthood. This allows us to evaluate what predicts the extent to which girls conform to gender norms, estimate the association between those norms and life cycle labor market outcomes, and investigate which channels mediate these results.

We find that conditional on a large set of age-11 individual characteristics, including cognitive and non-cognitive skills, girls who conform more strongly to gender norms earn less. This is due to both lower wages and fewer hours worked. Half of the association between earnings and gender conformity can be explained by selection into lower-paid occupations, lower educational attainment, and earlier marriage and child-bearing of those who conform more strongly with traditional gender roles. Our estimates on the determinants of gender conformity indicate that, conditional on a girl's skills, the wider environment in which they grow up (measured using region-level employment and educational attainment of women) shapes their gender conformity. We also find somewhat noisy estimates suggesting more gender conformity among girls with less educated parents and more female siblings.

Studying the determinants of gender conformity further, as well as disentangling the extent to which gender conformity is driven by heterogeneity in preferences versus constraints, are promising avenues for future work.

References

Adda, J., C. Dustmann, and K. Stevens (2017). The Career Costs of Children. Journal of Political Economy 125(2), 293–337.

Adukia, A., A. Eble, E. Harrison, H. B. Runesha, and T. Szasz (2023). What We Teach About Race and Gender: Representation in Images and Text of Children's Books. The Quarterly Journal of Economics 138(4), 2225–2285.

Agostinelli, F. and M. Wiswall (2016a). Estimating the Technology of Children's Skill Formation. Working Paper 22442, National Bureau of Economic Research.

Agostinelli, F. and M. Wiswall (2016b). Identification of Dynamic Latent Factor Models: The Implications of Re-Normalization in a Model of Child Development. Working Paper 22441, National Bureau of Economic Research.

Akerlof, G. A. and R. E. Kranton (2000). Economics and Identity. The Quarterly Journal of Economics 115(3), 715–753.

Alesina, A., P. Giuliano, and N. Nunn (2013). On the Origins of Gender Roles: Women and the Plough. The Quarterly Journal of Economics 128(2), 469–530.

Altonji, J. G., T. E. Elder, and C. R. Taber (2005). Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. Journal of Political Economy 113(1), 151–184.

Argamon, S., M. Koppel, J. Fine, and A. R. Shimoni (2003). Gender, Genre, and Writing Style in Formal Written Texts. Text & Talk 23(3), 321–346.

Ash, E., D. L. Chen, and A. Ornaghi (2024). Gender Attitudes in the Judiciary: Evidence from US Circuit Courts. American Economic Journal: Applied Economics 16(1), 314–350.

Banan, A. R., T. Santavirta, and M. Sarzosa (2023). Childhood Gender Nonconformity and Gender Gaps in Life Outcomes. Technical report, Working Paper.

Belfield, C., C. Crawford, E. Greaves, P. Gregg, and L. Macmillan (2017). Intergenerational Income Persistence within Families. Working Papers 17/11, IFS.

Ben-Porath, Y. (1967). The Production of Human Capital and the Life Cycle of Earnings. Journal of Political Economy 75(4, Part 1), 352–365.

Benjamin, D. J., J. J. Choi, and A. J. Strickland (2010). Social Identity and Preferences. American Economic Review 100(4), 1913–1928.

Bertrand, M. (2020). Gender in the Twenty-First Century. In AEA Papers and Proceedings, Volume 110, pp. 1–24. American Economic Association.

Bertrand, M., E. Kamenica, and J. Pan (2015). Gender Identity and Relative Income within Households. The Quarterly Journal of Economics 130(2), 571–614.

Bertrand, M. and S. Mullainathan (2001). Do people mean what they say? implications for subjective survey data. American Economic Review 91(2), 67–72.

Biasi, B. and S. Ma (2022). The Education-Innovation Gap. Working Paper 29853, National Bureau of Economic Research.

Black, N., D. W. Johnston, and A. Suziedelyte (2017). Justification Bias in Self-Reported Disability: New Evidence from Panel Data. Journal of Health Economics 54, 124–134.

Blanden, J., P. Gregg, and L. Macmillan (2013). Intergenerational Persistence in Income and Social Class: The Effect of Within-Group Inequality. Journal of the Royal Statistical Society: Series A (Statistics in Society) 176(2), 541–563.

Boelmann, B., A. Raute, and U. Schonberg (2021). Wind of Change? Cultural Determinants of Maternal Labor Supply.

Boerma, J. and L. Karabarbounis (2021). Inferring Inequality with Home Production. Econometrica 89(5), 2517–2556.

Bolt, U., E. French, J. H. Maccuish, and C. O'Dea (2021). The Intergenerational Elasticity of Earnings: Exploring the Mechanisms.

Bolukbasi, T., K.-W. Chang, J. Y. Zou, V. Saligrama, and A. T. Kalai (2016). Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. Advances in Neural Information Processing Systems 29.

Brenøe, A. A. (2022). Brothers Increase Women's Gender Conformity. Journal of Population Economics 35(4), 1859–1896.

Bursztyn, L., A. L. González, and D. Yanagizawa-Drott (2020). Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia. American Economic Review 110(10), 2997–3029.

Cagé, J., N. Hervé, and M.-L. Viaud (2020). The Production of Information in an Online World. The Review of Economic Studies 87(5), 2126–2164.

Cavapozzi, D., M. Francesconi, and C. Nicoletti (2021). The Impact of Gender Role Norms on Mothers' Labor Supply. Journal of Economic Behavior & Organization 186, 113–134.

Centre for Longitudinal Studies (2018). National Child Development Study: 'Imagine You Are 25' Essays (Sweep 2, Age 11). Data set.

Centre for Longitudinal Studies (2023a). National Child Development Study. Data Set.

Centre for Longitudinal Studies (2023b). National Child Development Study: Activity Histories, 1974-2013. Technical report, Institute of Education, University of London.

Cunha, F. and J. J. Heckman (2008). Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. Journal of Human Resources 43(4), 738–782.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. Econometrica 78(3), 883–931.

De Quidt, J., L. Vesterlund, and A. J. Wilson (2019). Experimenter Demand Effects. In Handbook of Research Methods and Applications in Experimental Economics, pp. 384–400. Edward Elgar Publishing.

Dhar, D., T. Jain, and S. Jayachandran (2019). Intergenerational Transmission of Gender Attitudes: Evidence from India. The Journal of Development Studies 55(12), 2572–2592.

Dhar, D., T. Jain, and S. Jayachandran (2022). Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India. American Economic Review 112(3), 899–927.

Downs, B., O. Anuyah, A. Shukla, J. A. Fails, S. Pera, K. Wright, and C. Kennington (2020). Kidspell: A Child-Oriented, Rule-Based, Phonetic Spellchecker. In Proceedings of the Twelfth Language Resources and Evaluation Conference, pp. 6937–6946.

Fernandez, R. (2007). Women, Work, and Culture. Journal of the European Economic Association 5(2-3), 305–332.

Fernández, R. and A. Fogli (2009). Culture: An Empirical Investigation of Beliefs, Work, and Fertility. American Economic Journal: Macroeconomics 1(1), 146–77.

Fernández, R., A. Fogli, and C. Olivetti (2004). Mothers and Sons: Preference Formation and Female Labor Force Dynamics. The Quarterly Journal of Economics 119(4), 1249–1299.

Field, E., R. Pande, N. Rigol, S. Schaner, and C. Troyer Moore (2021). On Her Own Account: How Strengthening Women's Financial Control Impacts Labor Supply and Gender Norms. American Economic Review 111(7), 2342–75.

Fortin, N. M. (2005). Gender Role Attitudes and the Labour-Market Outcomes of Women across OECD Countries. Oxford Review of Economic Policy 21(3), 416–438.

Gayle, G.-L. and L. Golan (2012). Estimating a Dynamic Adverse-Selection Model: Labour-Force Experience and the Changing Gender Earnings Gap 1968–1997. The Review of Economic Studies 79(1), 227–267.

Gayle, G.-L. and A. Shephard (2019). Optimal Taxation, Marriage, Home Production, and Family Labor Supply. Econometrica 87(1), 291–326.

Gelbach, J. B. (2016). When Do Covariates Matter? And Which Ones, and How Much? Journal of Labor Economics 34(2), 509–543. Publisher: University of Chicago.

Gentzkow, M., J. M. Shapiro, and M. Taddy (2019). Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech. Econometrica 87(4), 1307–1340.

Goldin, C. (2021). Career and Family: Women's Century-Long Journey toward Equity. Princeton University Press.

Goodman, A., M. Narayanan, S. Brown, and T. Murphy (2017). Age 11 Essays - Imagine You Are 25, 1969. NCDS user guide.

Goussé, M., N. Jacquemet, and J.-M. Robin (2017). Marriage, Labor Supply, and Home Production. Econometrica 85(6), 1873–1919.

Gregg, P. (2012). Occupational Coding for the National Child Development Study (1969, 1991-2008) and the 1970 British Cohort Study (1980, 2000-2008). Technical report, UCL.

Gregg, P., L. Macmillan, and C. Vittori (2016). Moving towards Estimating Sons' Lifetime Intergenerational Economic Mobility. Oxford Bulletin of Economics and Statistics.

Hanley, K. W. and G. Hoberg (2019). Dynamic Interpretation of Emerging Risks in the Financial Sector. The Review of Financial Studies 32(12), 4543–4603.

Hansen, S., M. McMahon, and A. Prat (2018). Transparency and Deliberation within the FOMC: A Computational Linguistics Approach. The Quarterly Journal of Economics 133(2), 801–870.

Heckman, J., R. Pinto, and P. Savelyev (2013). Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes. American Economic Review 103(6), 2052–86.

Heckman, J. J., J. Stixrud, and S. Urzua (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. Journal of Labor Economics 24(3), 411–482.

Jenkins, J. J., W. A. Russell, and G. J. Suci (1958). An Atlas of Semantic Profiles for 360 Words. The American Journal of Psychology 71(4), 688–699.

Kelly, B., D. Papanikolaou, A. Seru, and M. Taddy (2021). Measuring Technological Innovation over the Long Run. American Economic Review: Insights 3(3), 303–320.

Kern, M., A. Goodman, B. Wielgoszewska, M. Narayanan, J. Carpentieri, P. B. Ploubidis, George and, M. Zamani, and H. A. Schwartz (2022). Imagining the Future: Childhood Essays Predict Health, Cognitive, Economic, and Behavioural Outcomes Four Decades Later. Unpublished.

Kleven, H., C. Landais, and J. E. Søgaard (2019). Children and Gender Inequality: Evidence from Denmark. American Economic Journal: Applied Economics 11(4), 181–209.

Kozlowski, A. C., M. Taddy, and J. A. Evans (2019). The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings. American Sociological Review 84(5), 905–949.

Kozlowski, D. and I. Power (2022). Unpacking Backlash: Social Costs of Gender Non-Conformity for Women and Men. Journal of Research in Gender Studies 12(2).

Michalopoulos, S. and M. M. Xue (2021). Folklore. The Quarterly Journal of Economics 136(4), 1993–2046.

Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado, and J. Dean (2013). Distributed Representations of Words and Phrases and their Compositionality. Advances in Neural Information Processing Systems 26.

Nissim, M., R. van Noord, and R. van der Goot (2020). Fair Is Better than Sensational: Man Is to Doctor as Woman Is to Doctor. Computational Linguistics 46(2), 487–497.

Oster, E. (2019). Unobservable Selection and Coefficient Stability: Theory and Evidence. Journal of Business & Economic Statistics 37(2), 187–204.

Papageorge, N. W., V. Ronda, and Y. Zheng (2019). The Economic Value of Breaking Bad: Misbehavior, Schooling and the Labor Market. Technical report, National Bureau of Economic Research.

Patel, K. and P. Bhattacharyya (2017). Towards Lower Bounds on Number of Dimensions for Word Embeddings. In G. Kondrak and T. Watanabe (Eds.), Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers), Taipei, Taiwan, pp. 31–36. Asian Federation of Natural Language Processing.

Pennington, J., R. Socher, and C. D. Manning (2014). GloVe: Global Vectors for Word Representation. In Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543.

Pongiglione, B., M. L. Kern, J. Carpentieri, H. A. Schwartz, N. Gupta, and A. Goodman (2020). Do Children's Expectations about Future Physical Activity Predict Their Physical Activity in Adulthood? International Journal of Epidemiology 49(5), 1749–1758.

Sabra, S. and V. Sabeeh (2020). A Comparative Study of N-gram and Skip-gram for Clinical Concepts Extraction. In 2020 International Conference on Computational Science and Computational Intelligence (CSCI), pp. 807–812. IEEE.

Todd, P. E. and W. Zhang (2020). A Dynamic Model of Personality, Schooling, and Occupational Choice. Quantitative Economics 11(1), 231–275.

Vella, F. (1998). Estimating Models with Sample Selection Bias: A Survey. The Journal of Human Resources 33(1), 127.

Wielgoszewska, B., A. Goodman, M. Narayanan, M. Zamani, M. Kern, and H. A. Schwartz (2022). Linguistic Features of Childhood Essays can Predict Who Moves Up and Down the Income Ladder. Unpublished.

Wiswall, M. and B. Zafar (2018). Preference for the Workplace, Investment in Human Capital, and Gender. The Quarterly Journal of Economics 133(1), 457–507.

Wu, A. H. (2020). Gender Bias among Professionals: An Identity-Based Interpretation. Review of Economics and Statistics 102(5), 867–880.

Zhang, H., A. Sneyd, and M. Stevenson (2020). Robustness and Reliability of Gender Bias Assessment in Word Embeddings: The Role of Base Pairs. arXiv, 2010.02847. Preprint.

A Additional Results

A.1 Additional Tables

Table A.1: Full Lifetime Earnings Regression

	Log Lifetime Earnings	Log Lifetime Earnings	Log Lifetime Earnings
Gender Conformity Index	-0.055***	-0.029*	-0.035**
	(0.016)	(0.016)	(0.017)
Cognitive Skills		0.225***	0.201***
		(0.018)	(0.020)
Non-Cognitive Skills		0.026*	0.026
		(0.016)	(0.016)
Dad: Post Compulsory			0.053
			(0.046)
Dad: Some College			0.025
			(0.092)
Mum: Post Compulsory			-0.003
			(0.043)
Mum: Some College			0.165
			(0.118)
Log Parental Income			0.027
			(0.038)
Only Child			0.143*
			(0.078)
Sisters Only			0.034
			(0.054)
Sisters & Brothers			0.023
			(0.050)
Number of Siblings			0.014
			(0.014)
Birth Order			-0.042**
			(0.021)
Expect to go to School			0.070
			(0.096)
Expect to go to Uni			0.159
			(0.104)
Mother's Age			0.002
			(0.004)
Father's Age			-0.001
			(0.003)
Stable Marriage			-0.003
			(0.048)
Constant	12.223***	12.176***	11.671***
	(0.017)	(0.018)	(0.321)

Notes: Sample size is 3,497 girls. We report significance at the 1% (***), 5% (**), and 10% (*) levels, with robust standard errors. We present all coefficients (except dummies for missing regressors) in the lifetime earnings regressions; some coefficients from this regression are presented in Table 3 in Section 4.1.

Table A.2: Determinants of Gender Conformity Index

Determinant	Saturated Model	Univariate Models
Cognitive Skills	-0.083***	-0.119***
	(0.021)	(0.018)
Non-Cognitive Skills	-0.037**	-0.060***
	(0.017)	(0.017)
Log Parental Income	-0.016	-0.061
	(0.039)	(0.038)
Dad: Post Compulsory	-0.041	-0.137***
	(0.048)	(0.045)
Dad: Some College	-0.011	-0.190**
	(0.098)	(0.084)
Father's Age	-0.004*	-0.008***
	(0.002)	(0.002)
Mum: Post Compulsory	-0.032	-0.126***
	(0.044)	(0.040)
Mum: Some College	-0.085	-0.228**
	(0.113)	(0.109)
Mother's Age	-0.006	-0.011***
	(0.004)	(0.003)
Only Child	0.016	0.021
	(0.072)	(0.080)
Sisters Only	0.087	0.110*
	(0.055)	(0.057)
Sisters & Brothers	0.032	0.075
	(0.050)	(0.048)
Birth Order	0.008	-0.005
	(0.021)	(0.017)
Parent Exp: Uni	-0.051	-0.116***
	(0.037)	(0.038)
Mum's Hours Worked	-0.032	-0.010
	(0.030)	(0.018)
Mum Employed	0.052	0.014
	(0.061)	(0.035)
Female Headteacher	-0.034	-0.074*
	(0.045)	(0.042)
Women: Employment	-0.040**	-0.044**
	(0.018)	(0.019)
Women: Went to Uni	-0.076***	-0.078***
	(0.020)	(0.020)

Notes: Sample size is 3,497. We report significance at the 1% (***), 5% (**), and 10% (*) levels with robust standard errors. This table presents regression coefficients corresponding to those displayed in Figure 5. Column 1 are coefficients from a regression of the gender conformity index on all the listed determinants (the "saturated" model in Figure 5). Column 2 contains coefficients from a univariate regression of the index on the given determinant. We include dummies for missing regressor observations.

Table A.3: Shapley Values and Feature Importance

Determinant	Shapley Value	Shapley Rank	RF Regressor Importance	Feature Rank
Cognitive Skills	0.0071	1	1	1
Non-Cognitive Skills	0.0020	5	0.9326	2
Log Parental Income	0.0002	11	0.9281	3
Dad's Education	0.0010	9	0.5853	11
Father's Age	0.0021	3	0.8462	5
Mum's Education	0.0010	7	0.5334	13
Mother's Age	0.0020	4	0.8365	6
Sibling Composition	0.0008	10	0.6245	9
Birth Order	0.0001	13	0.6186	10
Parent Exp: Uni	0.0010	8	0.5478	12
Mum's Hours Worked	0.0002	12	0.7788	8
Mum Employed	0.0001	15	0.4759	15
Female Headteacher	0.0001	14	0.5078	14
Women: Employment	0.0013	6	0.8564	4
Women: Went to Uni	0.0042	2	0.7849	7

Notes: Sample size is 3,497. Column 1 reports Shapley values, decomposing the share of $R^2 = 0.035$ attributable to each variable in the saturated regression reported in Table A.2. Column 2 ranks determinants based on their Shapley value - a lower rank number indicates a larger contribution to the R^2. Column 3 reports regressor importance from a STATA random forest, with gender conformity as the outcome and the determinants in Table A.2 as the regressors. Variable importance is calculated by adding up the improvement in the objective function given in the splitting criterion over all internal nodes of a tree and across all trees in the forest, separately for each determinant variable. We use STATA's implementation of regressor importance, which is normalized by dividing all scores over the maximum score: The regressor importance of the most important variable is always 100%. Column 4 ranks determinants based on their importance in predicting the gender conformity index in the random forest. We include dummies for missings.

Table A.4: Intensive Margin of Labor Supply

	Log Total Hours	Log Total Hours	Log Total Hours
Gender Conformity Index	-0.013***	-0.010**	-0.010**
	(0.004)	(0.004)	(0.004)
Cognitive Skills		0.033***	0.031***
		(0.004)	(0.005)
Non-Cognitive Skills		0.001	-0.000
		(0.004)	(0.004)
Parental Education & Income			✓
Family Background			✓
County Fixed Effects			✓

Notes: Sample size is 3,497. We report significance at the 1% (***), 5% (**), and 10% (*) levels with robust standard errors. This table presents analogous results to those in Table 4 for the intensive margin of labor supply. Here, hours worked over the life cycle are computed conditional on employment status.

A.2 Robustness Checks

Table A.5: Instrumenting for Gender Conformity

	Lifetime Earnings	Lifetime Average Wages	Employed	Average Weekly Hours Worked
Gender Conformity Index	-0.036**	-0.019*	-0.005	-0.015*
	(0.018)	(0.011)	(0.004)	(0.009)
Cognitive Skills	0.200***	0.139***	0.015***	0.057***
	(0.020)	(0.013)	(0.004)	(0.010)
Non-Cognitive Skills	0.026	0.013	0.008**	0.010
	(0.016)	(0.010)	(0.004)	(0.008)
Parental Education & Income	✓	✓	✓	✓
Family Background	✓	✓	✓	✓
Geographic Fixed Effects	✓	✓	✓	✓

Notes: Sample size is 3,497 girls. We report significance at the 1% (***), 5% (**), and 10% (*) levels with robust standard errors. In this table, we instrument for the gender conformity index using 10 noisy measures of gender conformity from the text data. Each column reports results from regressions with the full set of controls: cognitive and non-cognitive skills, parental income and education, number of and sex composition of siblings, birth order, mother's and father's age at birth, parental aspirations, and family stability. We also include 111 county fixed effects. We include dummies for missing regressor observations.

Table A.6: Robustness to Choice of Controls

	Lifetime Earnings	Lifetime Average Wages	Employed	Average Weekly Hours Worked
Panel A: More Skills Controls				
Gender Conformity Index	-0.037**	-0.020*	-0.007	-0.018**
	(0.018)	(0.012)	(0.004)	(0.009)
Cognitive Skills	0.203***	0.137***	0.015**	0.055***
	(0.025)	(0.016)	(0.006)	(0.013)
Non-Cognitive Skills	0.022	0.012	0.006	0.008
	(0.018)	(0.012)	(0.004)	(0.008)
Panel B: School Quality Controls				
Gender Conformity Index	-0.035**	-0.020*	-0.005	-0.016*
	(0.017)	(0.011)	(0.004)	(0.008)
Cognitive Skills	0.200***	0.139***	0.015***	0.058***
	(0.021)	(0.013)	(0.005)	(0.010)
Non-Cognitive Skills	0.027	0.014	0.007**	0.009
	(0.016)	(0.011)	(0.004)	(0.008)
Panel C: Aspired Occupation				
Gender Conformity Index	-0.034*	-0.019*	-0.004	-0.014
	(0.017)	(0.011)	(0.004)	(0.009)
Cognitive Skills	0.201***	0.140***	0.015***	0.055***
	(0.021)	(0.014)	(0.005)	(0.010)
Non-Cognitive Skills	0.027*	0.013	0.008**	0.009
	(0.016)	(0.011)	(0.004)	(0.008)
Panel D: Region Fixed Effects				
Gender Conformity Index	-0.030*	-0.017	-0.005	-0.014*
	(0.016)	(0.010)	(0.004)	(0.008)
Cognitive Skills	0.200***	0.141***	0.015***	0.055***
	(0.020)	(0.013)	(0.004)	(0.010)
Non-Cognitive Skills	0.028*	0.013	0.008**	0.011
	(0.016)	(0.010)	(0.004)	(0.008)
Parental Education & Income	✓	✓	✓	✓
Family Background	✓	✓	✓	✓
County Fixed Effects	✓	✓	✓	✓

Notes: Sample size is 3,497. We report significance at the 1% (***), 5% (**), and 10% (*) levels with robust standard errors. Each column reports results from regressions with the full set of controls: cognitive and non-cognitive skills, parental income and education, number of and sex composition of siblings, birth order, mother's and father's age at birth, parental aspirations, and family stability. We also include 111 county fixed effects in Panels A, B and C. Each regression includes dummies for missing regressor observations. Panel A controls for quadratic skills variables, as well as essay length and number of spelling mistakes in the essay. Panel B adds 5 controls on school quality at age 11: type of school, class size, school size, student-teacher ratio, and percent of 11-year-olds in the cohort member's class who are ready for GCE exams. Panel C controls for the cohort member's aspired occupation. Panel D uses region, instead of county, fixed effects, since we observe the former at age 11 while the latter is constructed using age 16 county information (as discussed in Section 2).

Table A.7: Using Different Word Embedding Models

	Lifetime Earnings	Lifetime Average Wages	Employed	Average Weekly Hours Worked
Panel A: Pre-Trained Google Word2Vec WEM				
Gender Conformity Index	-0.018	-0.007	-0.003	-0.012
	(0.020)	(0.013)	(0.004)	(0.010)
Cognitive Skills	0.203***	0.141***	0.016***	0.058***
	(0.020)	(0.013)	(0.004)	(0.010)
Non-Cognitive Skills	0.028*	0.014	0.008**	0.010
	(0.016)	(0.011)	(0.004)	(0.008)
Panel B: Pre-Trained Stanford GloVe WEM				
Gender Conformity Index	-0.022	-0.010	-0.003	-0.014
	(0.021)	(0.013)	(0.004)	(0.010)
Cognitive Skills	0.203***	0.141***	0.016***	0.057***
	(0.020)	(0.013)	(0.004)	(0.010)
Non-Cognitive Skills	0.028*	0.014	0.008**	0.010
	(0.016)	(0.011)	(0.004)	(0.008)
Parental Education & Income	✓	✓	✓	✓
Family Background	✓	✓	✓	✓
County Fixed Effects	✓	✓	✓	✓

Notes: Sample size is 3,497. We report significance at the 1% (***), 5% (**), and 10% (*) levels with robust standard errors. Each column reports results from regressions with the full set of controls: cognitive and non-cognitive skills, parental income and education, number of and sex composition of siblings, birth order, mother's and father's age at birth, parental aspirations, and family stability. We also include 111 county fixed effects. We include dummies for missing regressors. Panels A & B construct the gender conformity index using Google's pre-trained Word2Vec WEM and Stanford's GloVe pre-trained WEM, respectively.

Table A.8: Index Construction

	Lifetime Earnings	Lifetime Average Wages	Employed	Average Weekly Hours Worked
Panel A: Occur ≥ 200 times, ≥ 2SD from neutral				
Gender Conformity Index	-0.035**	-0.021**	-0.005	-0.015*
	(0.016)	(0.010)	(0.004)	(0.008)
Panel B: Occur ≥ 200 times, include all neutral words				
Gender Conformity Index	-0.031*	-0.015	-0.005	-0.016*
	(0.016)	(0.011)	(0.004)	(0.008)
Panel C: Occur ≥ 10 times, ≥ 1SD from neutral				
Gender Conformity Index	-0.027	-0.017	-0.003	-0.012
	(0.017)	(0.015)	(0.004)	(0.009)
Panel D: Occur ≥ 100 times, ≥ 1SD from neutral				
Gender Conformity Index	-0.033*	-0.018*	-0.004	-0.013
	(0.017)	(0.011)	(0.004)	(0.009)
Panel E: Occur ≥ 300 times, ≥ 1SD from neutral				
Gender Conformity Index	-0.037**	-0.021*	-0.005	-0.015*
	(0.016)	(0.011)	(0.004)	(0.008)
Cognitive & Non-Cognitive Skills	✓	✓	✓	✓
Parental Education & Income	✓	✓	✓	✓
Family Background	✓	✓	✓	✓
County Fixed Effects	✓	✓	✓	✓

Notes: Sample size is 3,497. We report significance at the 1% (***), 5% (**), and 10% (*) levels with robust standard errors. Each column reports results from regressions with the full set of controls: cognitive and non-cognitive skills, parental income and education, number of and sex composition of siblings, birth order, mother's and father's age at birth, parental aspirations, and family stability. We also include 111 county fixed effects. Each regression also includes dummies for missing regressor observations. To construct our index in the main text, we use words that are used at least 200 times across all essays, and we drop all words that are within 1SD of being neutral on the gender dimension. This table shows how results change when we alter the cutoffs regarding (i) frequency of occurrence across essays and (ii) how close the gender score is to neutral. We only report coefficients on our gender conformity index. For our baseline results, 285 different words are used to construct the gender conformity index. From Panel A to Panel E, the number of word counts used to construct the index are: 67, 787, 1495, 423, and 228, respectively.

Table A.9: Results for Men: Lifetime Earnings, Wages, Hours, and Employment

	(1)	(2)	(3)
Panel A: Log Lifetime Earnings			
Sample mean: 13.08			
Gender Conformity Index	-0.037**	0.003	0.003
	(0.018)	(0.018)	(0.018)
Cognitive Skills		0.233***	0.191***
		(0.013)	(0.015)
Non-Cognitive Skills		0.009	0.013
		(0.013)	(0.013)
Panel B: Log Lifetime Average Hourly Wages			
Sample mean: 1.99			
Gender Conformity Index	-0.033**	0.005	0.004
	(0.016)	(0.016)	(0.016)
Cognitive Skills		0.224***	0.191***
		(0.011)	(0.013)
Non-Cognitive Skills		0.003	0.004
		(0.011)	(0.011)
Panel C: Log Total Lifetime Hours Worked			
Sample mean: 11.02			
Gender Conformity Index	-0.009*	-0.003	0.000
	(0.005)	(0.005)	(0.005)
Cognitive Skills		0.034***	0.024***
		(0.005)	(0.005)
Non-Cognitive Skills		0.010**	0.013***
		(0.004)	(0.004)
Panel D: Share of Lifetime Spent in Employment			
Sample mean: 0.94			
Gender Conformity Index	-0.005	-0.002	0.000
	(0.003)	(0.003)	(0.003)
Cognitive Skills		0.017***	0.012***
		(0.002)	(0.003)
Non-Cognitive Skills		0.006**	0.009***
		(0.002)	(0.002)
Parental Education & Income			✓
Family Background			✓
County Fixed Effects			✓

Notes: Sample size is 3,513 boys. We report significance at the 1% (***), 5% (**), and 10% (*) levels with robust standard errors. The first column shows the univariate OLS regression of each lifetime outcome on our gender conformity index. Column (2) includes controls for cognitive and non-cognitive skills. Column (3) includes our full set of family background and geographic controls, including (as displayed) parental income and education. We also control for (but do not display): number of and sex composition of siblings, birth order, mother's and father's age at birth, parental aspirations, and family stability. We include 111 county fixed effects. All specifications have dummies for missing regressor observations.

B Data

B.1 Sample Selection

Table B.1 illustrates our sample selection criteria in the order in which we apply them. The full NCDS sample is 18,558 individuals born in the third week of March in 1958. 15,335 of these individuals are surveyed at age 11, of which 13,732 cohort members wrote essays. 10,551 essays were successfully digitized, of which 5,091 are female.[31]

Since lifetime earnings is our main outcome, we first drop the 1,544 individuals for whom we do not observe earnings at any age between 23 and 55 (inclusive). This is the case if there is no information on earnings in these five survey waves, if the individual is never employed at the time of the survey, or if we observe less than 60 months of working history for them. We further exclude two individuals for whom we do not have any recorded information on hours worked in their activity history (Centre for Longitudinal Studies, 2023b). Finally, we drop all remaining non-white girls to arrive at our final sample of 3,497.

Table B.1: Sample Selection Criteria

Sample Selection Criteria	# of Individuals
Total Sample in 1958	18,558
Observed at Age 11	15,335
Wrote Essays at Age 11	10,551
Sample of Girls' Essays	5,091
Observe earnings / wages at least once	3,547
Observe hours worked at least once	3,545
Exclude non-white girls	3,497
Final Sample	**3,497**

Notes: Here, we present our sample selection criteria, and the number of individuals who remain after cumulatively applying our criteria. For instance, 10,551 is the number of individuals in the sample if we take the total sample in 1958 who are observed at age 11 *and* who wrote essays at age 11.

Table B.2 affirms that our main results are robust to the specific sample selection criteria that we impose in our main specifications.

B.2 Measuring Cognitive and Non-Cognitive Skills

We assume that cognitive and non-cognitive skills at age 11 are not measured directly; instead we have noisy measures of underlying latent factors. Following previous literature

[31] Goodman et al. (2017) describe that some essays could not be digitized due to the microfiche being missing or illegible.

Table B.2: Robustness to Sample Selection

	Lifetime Earnings	Lifetime Average Wages	Employed	Average Weekly Hours Worked
Panel A: Include Non-White Girls				
Gender Conformity Index	-0.036**	-0.021**	-0.005	-0.015*
	(0.016)	(0.011)	(0.004)	(0.008)
Cognitive Skills	0.203***	0.138***	0.017***	0.060***
	(0.020)	(0.013)	(0.004)	(0.010)
Non-Cognitive Skills	0.024	0.012	0.007*	0.008
	(0.016)	(0.010)	(0.004)	(0.008)
Number of Observations	3545	3545	3545	3545
Panel B: No Exclusion Criteria (Among 5,091 Girls Essays)				
Gender Conformity Index	-0.036**	-0.021**	-0.005	-0.015*
	(0.016)	(0.011)	(0.004)	(0.008)
Cognitive Skills	0.203***	0.138***	0.017***	0.060***
	(0.020)	(0.013)	(0.004)	(0.010)
Non-Cognitive Skills	0.024	0.012	0.007**	0.008
	(0.016)	(0.010)	(0.004)	(0.008)
Number of Observations	3547	3547	3549	3545
Parental Education & Income	✓	✓	✓	✓
Family Background	✓	✓	✓	✓
County Fixed Effects	✓	✓	✓	✓

Notes: We report significance at the 1% (***), 5% (**) and 10% (*) levels with robust standard errors. Each column reports results from regressions with the full set of controls: cognitive and non-cognitive skills, parental income and education, number of and sex composition of siblings, birth order, mother's and father's age at birth, parental aspirations and family stability. We also include 111 county fixed effects. Each regression includes dummies for missing observations. Each panel uses different sample selection criteria.

(Cunha and Heckman, 2008; Agostinelli and Wiswall, 2016b), we assume a linear relationship between measures (Z) and underlying latent factors $\omega \in \{C, NC\}$:

$$Z_{\omega,i,j} = \mu_{\omega,j} + \lambda_{\omega,j}\omega_i + \epsilon_{\omega,i,j}$$

Here, $Z_{\omega,i,j}$ denotes measure j of latent factor ω (e.g. a math score as a measure of latent cognitive skills) for individual i at age 11. $\mu_{\omega,j}$ and $\lambda_{\omega,j}$, respectively, are the location and scale of this measure and are constant across individuals. $\epsilon_{\omega,i,j}$ denotes an idiosyncratic measurement error, assumed to be independent across individuals, measures, and time. The measurement errors are also assumed to be independent of the latent variables and all other controls and shocks. As the latent factors do not have a natural scale or location, we

normalize their means to be zero in every period and their variances to be one. This allows us to estimate the location and scale parameter for each measure (see Bolt et al. (2021) for details). We then predict the latent factors for each individual, using the Bartlett score method (Heckman et al., 2013). Bartlett scores are a linear combination of all measures used, inversely weighted by their noise. This means that measures with little measurement error get more weight than those with a lot of measurement error, thus minimizing measurement error in the final score. These Bartlett scores are the indices of cognitive and non-cognitive skills that we use in all our regression specifications.

The measures we use for cognitive skills are: math scores, reading scores, and a test which asked cohort members to copy designs across worksheets, all measured at age 11. The measures we use for non-cognitive skills are mother-reported items regarding whether the child is unconcentrated, is solitary, gets bullied, is destructive, appears miserable, is fidgety, gets worried easily, is irritable, sucks on his/her thumb, gets upset easily, has mannerisms, fights frequently, bites nails, and is disobedient.

B.3 Parental Income

Parental income variables are observed in the third wave of the NCDS, that is in 1974, when cohort members are aged 16. There are three relevant variables: fathers' net earnings, mothers' net earnings, and other net income to the household.

Fathers' and mothers' earnings and other net income are recorded as interval, or "banded" data. We use the continuous parental income variables which were imputed by the Institute for Fiscal Studies as part of an effort to harmonize income variables in different cohort studies (Belfield et al., 2017). This procedure can be summarized by three steps. First, income bands identical to the NCDS are created for the 1973, 1974, and 1975 Family Expenditure Survey (FES) data. Second, within each income band, net male, net female, and net other income data are estimated using variables that are correlated with income.[32] Third, using these prediction equations, each of the three income components are separately predicted within each income band in the NCDS, using the same covariates. The sum of the three net, predicted earnings variables in the NCDS is our measure of parental income.

[32]The following variables are used to estimate income data within each band: year of interview, age left school of mother and father, employment status of mother and father, age of mother and father, occupation of mother and father, number of siblings of the cohort member, housing tenure, region, number of rooms in household, number of cars, marital status, whether benefits are received by household, type of school cohort member attends, interactions between age left school and occupation of mother and father, interactions between age left school and employment status of mother and father, and interactions between year of interview and occupation of mother and father

For parents who did not answer any of the earnings questions, parental income is set as missing. Adopting Blanden et al. (2013)'s procedure, we also set parental income to missing where at least one parent is working but that parent's earnings are unobserved.

B.4 Earnings and Wages of Cohort Members

NCDS cohort members' gross wages and hours worked are observed at ages 23, 33, 42, 50, and 55. Gross earnings are calculated using usual gross wages and pay period variables on the respondent's main job. Current or last wages are used where usual wages are unavailable.

The crucial step is in how we address missing and zero wages and earnings figures for cohort members, which could be due to attrition, item non-response or non-employment. To address this, we adapt Gregg et al. (2016)'s method of imputing missing wages/earnings at the time of a survey. We model wages/earnings as a function of an individual-specific fixed effect and a quadratic in age:

$$Y_{it} = f_i + \chi_1 \cdot time_t + \chi_2 \cdot (yos_{it} * time_t) + \zeta_{it}$$

where Y_{it} is wages or earnings, f_i is the individual fixed effect and yos_i is individual i's years of schooling and $time_t$ is the (common) age of NCDS cohort members at time t. Earnings and wages at ages 23, 33, 42, 50, and 55 are predicted as $\widehat{Y_{it}} = \hat{f}_i + \hat{\chi}_1 \cdot time_t + \hat{\chi}_2 \cdot (yos_{it} * time_t)$.

Earnings for individuals recorded as being out of work at the time of a survey is set to zero since missing earnings in those cases represents non-employment as verified in the more detailed NCDS Activity History data (Centre for Longitudinal Studies, 2023b). We also set earnings to zero whenever $\widehat{Y_{it}} < 0$. Monthly earnings for every month are calculated from earnings at each survey date by imputing a linear trajectory between each earnings data point for every month between surveys. Aggregating monthly earnings provides our estimate of cohort members' lifetime earnings.

C Estimating Gender Conformity

C.1 Steps 1 and 2: Spelling Correction and Pre-Processing

The original NCDS essays from 1969 were handwritten. In 2018, the Centre for Longitudinal Studies (CLS) digitized the essays, redacting only names and other identifying information.

To be able to estimate gender conformity and ensure it does not pick up on latent cognitive skills, we correct spelling mistakes in the essays. The procedure is as follows.

1. The essays are broken up into individual sentences, based on periods used.

2. For each sentence, misspelled words are identified using the UK English dictionary in Python's *pyenchant* library. A sentence, along with a list of its misspelled words, are outputted to Microsoft Excel – sentences from the same essay are in the same sheet.

3. A team of 15 undergraduate students manually correct words that are identified as misspelled. The precise instructions are given in Figure 6. When making these corrections, each student can infer the word's use in the context of the sentence, and in principle, in the context of the other sentences of the essay.

Figure 6: Instructions for Spelling Correction Task

Each excel file contains 25 excel sheets. Each excel sheet contains some sentences from an essay, and each of these sentences contains one or more misspelled words. In column P, the full sentence is given and columns E to O contain misspelled words. The objective is to identify the correct spelling of the word and enter it in the cell directly beneath the misspelled words listed in Columns E to O. Please do not edit the sentences or any other information on the excel sheets.

Refer to the sample file "set25.xslx" for what we would give to you and "set25_corrected.xslx" for what we would hope to get back.

There are certain specific cases which require explanation:

1. **If you see "xxx" or "***" as a misspelled word:**
   a. If followed by a number (e.g. "xxx120") it often refers to "£" – if this fits the context of the sentence, simply replace "xxx120" with "120 pounds" (see sheet essay7)
   b. If the sequence of x's or *'s does not refer to anything in particular, simply re-enter the sequence of x's or *'s in the cell underneath (see sheet essay20)
2. **If you are not sure but think you know the answer**
   a. Write in your guess – if you do this, enter the guess in the cell directly underneath the misspelled word, and enter "guess!" directly beneath your guess (see essay9, cell F17)
3. **If you cannot decipher the word:**
   a. enter "idk!" directly beneath the misspelled word (see cell F25 in sheet essay5)
4. **Some words are "split" in the original essays** –
   a. Some words may be split, e.g. in essay24 "Alsatian" is spelt "al sation", and "al" and "sation" are separately flagged as misspelled. In this case, enter the correct word "Alsatian" under one of these two words, and leave the other blank.
5. **Abbreviations are sometimes given as misspelled words:**
   a. e.g. B.U. in essay 5. In this instance, re-enter the abbreviation under the flagged word

Points of Information:

- Avoid symbols where possible – use "pounds" instead of "£"
- Avoid short-forms where possible – if "cm" is flagged, write "centimetres"
- For context, these essays were written in 1969/1970. The currency at the time was pounds, shillings and pence. E.g. in essay4, "xxx12 10s 6d" stands for 12 pounds, 10 shillings, 6 pence.

4. Misspelled words are automatically replaced with correctly spelled ones, and sentences are merged together to reconstruct the essays. These are our final essay data.

The final essays are then read into Python as a matrix of 1-gram counts using Python's *sklearn* library. All stop words, which are pre-defined in the library, are removed in this

process. In addition, words which appear less than 200 times across all essays are removed. This is our bag-of-words matrix.Then, 2-grams starting with "not" or "no" are read in. We append counts of 2-grams if their first word is among those in our bag-of-words matrix, and if the second word of the 2-gram is not a stop word. Thus, we arrive at our final representation for the essays.

C.2 Step 3: Training our Word-Embedding Model

We assign vector values to two groups of words. The first is words in our Bag-of-Words matrix of counts. The second is the ten pairs of words we use to construct our gender dimension. To assign these vector values, we train our own Word-Embedding Model using Google Books Ngrams data.[33] This dataset includes all n-grams that appear over 40 times across a corpus consisting of about 8 million books. We restrict the period in which the texts were written to be between 1958 and 1978, to ensure our model captures word meanings that children learn and use in their age 11 essays. The amount of text data available for this period is relatively small compared to the entire corpus of 8 million books – we use 1-,2-,3-,4- and 5-grams found in 2.6 million books to train our model. We use the *genism* library in Python and adapt code from Kozlowski et al. (2019) to train the Word-embedding model.[34]

The specifications of our training process are given as below:

- Algorithm = Skip-Gram
- Vector size = 300
- Window = 5 (max. distance between the current & predicted word within a sentence)
- Min Count = 10 (ignore all words with total frequency lower than this)
- Negative sampling used. 8 "noise" words are drawn per iteration.

The model is trained on a 36-core cluster and took around 80 hours to train.

C.3 Validating our Procedure

C.3.1 Gender Profiles of Words

Jenkins et al. (1958) ask 540 psychology students to rate the gender connotations of 360 words and phrases[35]. Participants had to assign a number between 1 and 7 to each word, where a higher rating meant they considered the word more feminine.

[33] storage.googleapis.com/books/ngrams/books/datasetsv3.html
[34] github.com/KnowledgeLab/GeometryofCulture
[35] Of these, we focus on the 353 1-grams.

Following Kozlowski et al. (2019), we use Jenkins et al. (1958)'s data to validate the projection values that our word-embedding model (WEM) assigns to words. To do so, we first find the rank correlation between the projection values our WEM assigns to 353 words, and the corresponding average rating for each word assigned by participants of Jenkins et al. (1958)'s study. Reassuringly, this turns out to be 0.52, suggesting that our model ranks words in terms of their gender connotation in a comparable way to 540 young adults, a few years before the NCDS essays were written. A simple regression of the projection values from our WEM on the human-assigned scores in Jenkins et al. (1958)'s study yields an R^2 of 27%, suggesting that our WEM procedure captures a larger amount of variation in the gender connotation of words – Kozlowski et al. (2019) discuss (in Appendix G) that this could be due to the averaging of ratings in Jenkins et al. (1958).

C.3.2 Simpler Indices

We construct a simple index based on word counts of masculine/feminine words in the essays. Comparing this simple index with our gender conformity index not only highlights the advantages of the gender conformity index (as discussed in Section 3.4), but also justifies that our estimation procedure draws on the variation in the essays as we would expect.

We selected the following words that all of us agreed to traditionally have strong gender connotations: husband, house, boy, girl, little, baby, stay, cook, wash, clean, breakfast, knit, tidy, kitchen, housework, sew, hostess, wedding, washing, part, and let. For each essay, we count how many times each of these words are used, and residualize this count on a quadratic of number of words in an essay. This is our naive index of gender conformity.

The first row in Table C.1 reports a raw correlation of 0.46 between the naive index and our preferred index of gender conformity. In a simple regression of our index on the naive one, we find an R^2 of 0.21; simple counts of the aforementioned words only explain 21% of the variation in our index. This could be for two reasons. Firstly, our procedure weights word counts by how gendered a given word is. Secondly, our preferred estimation does not specify a list of gendered words; in principle, any word can carry gender connotations.

To understand which of these two reasons dominate, the second row in Table C.1 repeats the exercise in the first row, but this time weights word counts in the naive index by their gender association, as given by our Word-Embedding Model. Both the R^2 and raw correlation are unchanged, suggesting that the unexplained variation in our preferred index comes from accounting for the gender connotation of *all* words, as opposed to a small subset.

Table C.1: Correlation of the Gender Conformity Index with Alternative Simpler Indices of Gender Conformity

	Raw Correlation	R^2 from Simple Regression
Naive Index	0.462	0.213
Naive Weighted Index	0.460	0.212

Notes: Correlations and R^2 are computed for analysis sample of 3,497 girls. The first column reports the raw correlations between the naive and naive weighted index with our preferred index of gender conformity. The second column reports the R^2 from a regression of our index both naive indices.

C.4 Other Latent Dimensions

In Section 4.4, we show robustness of our main results to the inclusion of six indices capturing further content of the essays. Following Kozlowski et al. (2019), the dimensions we use to construct these indices are: (1) education (educated-uneducated), (2) cultivation (cultured-uncultured), (3) affluence (rich-poor), (4) status (prestigious-undistinguished), (5) morality (good-evil), and (6) employment (employer-employee). Below, we list the pairs of antonyms used to construct each of these dimensions.

Education educated-uneducated, learned-unlearned, knowledgeable-ignorant, trained-untrained, taught-untaught, literate-illiterate, schooled-unschooled, tutored-untutored, lettered-unlettered

Cultivation cultivated-uncultivated, cultured-uncultured, civilized-uncivilized, courteous-uncourteous, proper-improper, polite-rude, cordial-uncordial, formal-informal, courtly-uncourtly, urbane-boorish, polished-unpolished, refined-unrefined, civility-incivility, polite-blunt urbanity-boorishness, politesse-rudeness, edified-loutish, mannerly-unmannerly, polished-gruff, gracious-ungracious, obliging-unobliging, cultured-uncultured, genteel-ungenteel, mannered-unmannered

Affluence rich-poor, richer-poorer, richest-poorest , affluence-poverty, affluent-destitute, advantaged-needy, wealthy-impoverished, costly-economical, exorbitant-impecunious, expensive-inexpensive, exquisite-ruined, sumptuous-plain, extravagant-necessitous, swanky-basic, flush-skint, thriving-disadvantaged, invaluable-cheap, upscale-squalid, lavish-economical, valuable-valueless, luxuriant-penurious, classy-beggarly, luxurious-threadbare, ritzy-ramshackle, luxury-cheap, opulence-indigence, solvent-insolvent, opulent-indigent, moneyed-moneyless, plush-threadbare, rich-penniless, luxuriant-penurious, affluence-penury, posh-plain, opulence-indigence,

precious-cheap, priceless-worthless, privileged-underprivileged, propertied-bankrupt, developed-underdeveloped, solvency-insolvency, successful-unsuccessful

Status honorable-dishonorable, esteemed-lowly, influential-uninfluential, reputable-disreputable, distinguished-commonplace, eminent-mundane, illustrious-humble, renowned-prosaic, acclaimed-modest, venerable-unpretentious, exalted-ordinary, estimable-lowly, prominent-common

Morality good-evil, moral-immoral, good-bad, honest-dishonest, virtuous-sinful, virtue-vice, righteous-wicked, chaste-transgressive, principled-unprincipled, unquestionable-questionable, noble-nefarious, uncorrupt-corrupt, scrupulous-unscrupulous, altruistic-selfish, chivalrous-knavish, honest-crooked, commendable-reprehensible, pure-impure, dignified-undignified, holy-unholy, valiant-fiendish, upstanding-villainous, guiltless-guilty, decent-indecent, chaste-unsavory, righteous-odious, ethical-unethical

Employment employer-employee, employers-employees, owner-worker, owners-worker, industrialist-laborer, industrialists-laborers, proprietor-employee, proprietors-employees, capitalist-proletarian, capitalists-proletariat, manager-staff, director-employee, directors-employees, boss-worker, bosses-workers, foreman-laborer, supervisor-staff, superintendent-staff

D Gelbach Decomposition

To quantify how much of the association between our gender conformity index and outcomes are attributable to family formation, education and occupation, we use the decomposition introduced by Gelbach (2016). In our application, this involves comparing the coefficient on the gender conformity index in our saturated regressions (equation (1) in the main text, reproduced in equation (5) below) to the coefficient on the gender conformity index in a regression which additionally includes our mediators (equation (6)):

$$y_i = \beta_0 + \beta_1 \text{GCI}_i + \beta_2 \text{Cog11}_i + \beta_3 \text{NonCog11}_i + \gamma' \mathbf{X_i} + u_i \tag{5}$$

$$y_i = \tilde{\beta}_0 + \tilde{\beta}_1 \text{GCI}_i + \tilde{\beta}_2 \text{Cog11}_i + \tilde{\beta}_3 \text{NonCog11}_i + \tilde{\gamma}' \mathbf{X_i} + \tilde{\delta}' \mathbf{M}_i + e_i \tag{6}$$

where y_i is the outcome of interest, GCI is our gender conformity index, Cog_{11} and Non-Cog_{11} are cognitive and non-cognitive skills at 11, $X_{i,t}$ is the vector of covariates that are included in our saturated regression models and M_i is the vector of our mediating variables (years of schooling, marital status and number children at 23, and occupation).

The difference between the two coefficients on the gender conformity index, $\beta_1 - \tilde{\beta}_1$, is the portion of association between gender conformity index and an outcome, attributable to

the mediators. The Gelbach decomposition allows us to allocate this difference among each of the mediators by relying on the formula for omitted variable bias:

$$\beta_1 - \tilde{\beta}_1 = \sum\nolimits_{j \in M} \kappa_{M_j}^{GCI} \tilde{\delta}_{M_j},$$

where $\kappa_{M_j}^{GCI}$ is the coefficient from a univariate regression of the jth element of $\mathbf{M}_i$ on the gender conformity index and $\tilde{\delta}_{M_j}$ is the jth element of the coefficient vector $\tilde{\delta}$. The share of the association between gender conformity index and the outcome explained by the jth mediator is thus given by $\kappa_{M_j}^{GCI}\tilde{\delta}_{M_j}/(\beta_1 - \tilde{\beta}_1)$. Because all the mediators are added simultaneously, the Gelbach decomposition is invariant to the order in which the factors enter.

Note, we include missing dummies in all equations. In addition to the missing dummies in equation (5), equation (6) additionally contains missing dummies for the mediators. Because these missing dummies explain a small fraction of the association between the gender conformity index and outcomes, the total share explained by education, family formation and occupation in the last column of Table 7 does not exactly equal the difference $\beta_1 - \tilde{\beta}_1$.

E Model

As described in Section 5.2, the maximization problem in adulthood is:

$$\begin{aligned} V_2(ed;\gamma,\xi) &= \max_{c_2,l_2,hp_2} u(c_2,l_2,hp_2;\gamma,\xi) \\ &= \max_{c_2,l_2,hp_2} \frac{1}{\tau}\ln[(1-\gamma)c_2^{\tau} + \gamma \cdot hp_2^{\tau}] + \delta \ln(l_2) \end{aligned}$$

subject to:

$$\begin{aligned} T &= hp_2 + l_2 + h_2 \\ c_2 &= h_2 \cdot w_2(ed) \end{aligned}$$

Writing this constrained maximisation in Lagrangean form:

$$\begin{aligned} L &= \max_{c_2,l_2,hp_2} u(c_2,l_2,hp_2;\gamma,\xi) + \mu[h_2 w_2(ed) - c_2] \\ &= \max_{c_2,l_2,hp_2} \frac{1}{\tau}\ln[(1-\gamma)c_2^{\tau} + \gamma \cdot hp_2^{\tau}] + \delta \ln(l_2) + \mu[(T - hp_2 - l_2)\cdot w_2(ed) - c_2] \end{aligned}$$

We derive the following first-order conditions:

$$\frac{(1-\gamma)c_2^{\tau-1}}{(1-\gamma)c_2^{\tau}+\gamma\cdot hp_2^{\tau}}=\mu$$

$$\frac{\gamma(hp_2)^{\tau-1}}{(1-\gamma)c_2^{\tau}+\gamma(hp_2)^{\tau}}=\mu w_2(ed)$$

$$\frac{\delta}{l_2}=\mu w_2(ed)$$

Combining these conditions gives:

$$\frac{\gamma hp_2^{\tau-1}}{(1-\gamma)c_2^{\tau}+\gamma hp_2^{\tau}}=\frac{\delta}{l_2}$$

$$\frac{w_2(ed)(1-\gamma)c_2^{\tau-1}}{(1-\gamma)c_2^{\tau}+\gamma hp_2^{\tau}}=\frac{\delta}{l_2}$$

which we can write as:

$$\begin{aligned}
\left(\frac{hp_2}{c_2}\right)^{\tau-1} &= \frac{w_2(ed)(1-\gamma)}{\gamma} \\
\frac{c_2}{hp_2} &= \left(\frac{w_2(ed)(1-\gamma)}{\gamma}\right)^{\frac{1}{1-\tau}} \\
\frac{h_2 w_2(ed)}{hp_2} &= w_2(ed)^{\frac{1}{1-\tau}}\left(\frac{(1-\gamma)}{\gamma}\right)^{\frac{1}{1-\tau}} \\
\frac{h_2}{hp_2} &= w_2(ed)^{\frac{\tau}{1-\tau}}\left(\frac{(1-\gamma)}{\gamma}\right)^{\frac{1}{1-\tau}} \qquad (7)
\end{aligned}$$

which is equation (4) in the main text. Using the time constraint, $h_2 = T - hp_2 - l_2$, we can also rewrite this as:

$$\frac{(T-hp_2-l_2)}{hp_2} = w_2(ed)^{\frac{\tau}{1-\tau}}\left(\frac{(1-\gamma)}{\gamma}\right)^{\frac{1}{1-\tau}} \qquad (8)$$

Turning to period 1 (youth), we reproduce the maximization problem in Section 5.2:

$$\begin{aligned}
V_1(\gamma,\xi) &= \max_{ed,c_1,l_1} u(c_1,l_1)+\beta V_2(ed;\gamma,\xi) \\
&= \max_{ed,c_1,l_1} \ln(c_1)+\delta\ln(l_1)+\beta V_2(ed;\gamma,\xi)
\end{aligned}$$

subject to

$$\begin{aligned} T &= l_1 + \xi ed + h_1 \\ c_1 &= h_1 \cdot w_1 \end{aligned}$$

Solving out for consumption, the maximization problem can be rewritten as

$$V_1(\gamma, \xi) = \max_{ed, l_1} \ln((T - l_1 - \xi ed)w_1) + \delta \ln(l_1) + \beta V_2(ed; \gamma, \xi)$$

We derive the following first-order conditions:

$$FOC_{l_1} : -\frac{w_1}{c_1} + \frac{\delta}{l_1} = 0$$

$$FOC_{ed} : \frac{-\xi w_1}{c_1} + \beta \frac{\partial V_2(ed; \gamma, \xi)}{\partial ed} = 0$$

The second first order condition can be rewritten as:

$$\beta \frac{\partial V_2(ed; \gamma, \xi)}{\partial ed} = \frac{\partial u(c_1, l_1)}{\partial c_1} \xi w_1 \tag{9}$$

which says that the marginal benefit of education (through higher wages in adulthood) equals the marginal cost today through lost wages and also equals the marginal benefit of working.

Now, recall that $V_2(ed; \gamma, \xi) = \frac{1}{\tau} \ln[(1-\gamma)((T - hp^* - l_2^*) \cdot w_2(ed))^\tau + \gamma(hp_2^*)^\tau] + \delta \ln(l_2^*)$, where we denote optimal levels chosen in adulthood with an asterisk. Thus,

$$\frac{\partial V_2(ed; \gamma, \xi)}{\partial ed} = \frac{(1-\gamma)(h_2^*)^\tau \cdot w_2(ed)^{\tau-1} w_2'(ed)}{(1-\gamma)(c_2^*)^\tau + \gamma(hp_2^*)^\tau}$$

Thus,

$$\frac{\partial V_2(ed; \gamma, \xi)}{\partial ed} = \frac{(1-\gamma)(T - hp_2^* - l_2^*)^\tau \cdot w_2(ed)^{\tau-1} w_2'(ed)}{(1-\gamma)(c_2^*)^\tau + \gamma(hp_2^*)^\tau} = \frac{(T - hp_2^* - l_2^*)^\tau \cdot w_2(ed)^{\tau-1} w_2'(ed)}{(c_2^*)^\tau + (\gamma/(1-\gamma))(hp_2^*)^\tau}$$

Using this equation and noting that $c_1 = w_1 h_1$ and $c_2 = w_2 h_2$, we rewrite Equation (9) as

$$\begin{aligned}
\beta \frac{(T - hp_2^* - l_2^*)^\tau \cdot w_2(ed)^{\tau-1}}{((T - hp_2^* - l_2^*) \cdot w_2(ed))^\tau + \frac{\gamma}{1-\gamma}(hp_2^*)^\tau} w_2'(ed) &= \frac{\xi w_1}{c_1} \\
\Rightarrow \beta \left[(T - hp_2^* - l_2^*)^\tau \cdot w_2(ed)^{\tau-1} w_2'(ed) \right] \frac{1}{\xi}(T - l_1 - \xi ed) &= ((T - hp_2^* - l_2^*) \cdot w_2(ed))^\tau + \frac{\gamma}{1-\gamma}(hp_2^*)^\tau \\
w_2(ed)^{\tau-1}(T - hp_2^* - l_2^*)^\tau \left[\beta \frac{1}{\xi}(T - l_1 - \xi ed) - w_2(ed) \right] &= \frac{\gamma}{1-\gamma}(hp_2^*)^\tau \\
w_2(ed)^{\tau-1}(h_2^*)^\tau \left[\beta \frac{1}{\xi} h_1 w_2'(ed) - w_2(ed) \right] &= \frac{\gamma}{1-\gamma}(hp_2^*)^\tau
\end{aligned}$$

Using Equation (8), we can simply this expression further:

$$\begin{aligned}
\frac{\gamma}{1-\gamma} &= w_2(ed)^{\tau-1}(h_2^*/hp_2^*)^\tau \left[\beta \frac{1}{\xi} h_1 w_2'(ed) - w(ed) \right] \\
&= w_2(ed)^{\tau-1} \left(w_2(ed)^{\frac{\tau}{1-\tau}} \left(\frac{(1-\gamma)}{\gamma} \right)^{\frac{1}{1-\tau}} \right)^\tau \left[\beta \frac{1}{\xi} h_1 w_2'(ed) - w(ed) \right] \\
&= w_2(ed)^{\frac{2\tau-1}{1-\tau}} \left(\frac{(1-\gamma)}{\gamma} \right)^{\frac{\tau}{1-\tau}} \left[\beta \frac{1}{\xi} h_1 w_2'(ed) - w(ed) \right] \\
\Rightarrow \left(\frac{\gamma}{1-\gamma} \right)^{\frac{1}{1-\tau}} &= w_2(ed)^{\frac{2\tau-1}{1-\tau}} \left[\beta \frac{1}{\xi} h_1 w_2'(ed) - w_2(ed) \right]
\end{aligned}$$

which is Equation (3) in Section 5.2.